

Best Available Copy

AD-A282 492



2

DTIC
ELECTE
JUL 25 1994
S G D

Modeling and Forecasting the Demand for Aircraft Recoverable Spare Parts

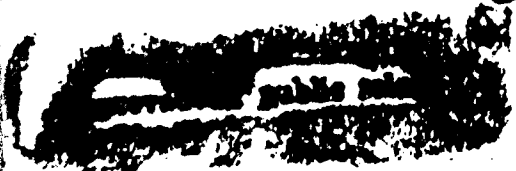
John L. Adams, John B. Abell, Karen E. Isaacson

copy

94-22919



94 7 21 036



DTIC QUALITY ASSURED

RAND

PROJECT AIR FORCE
NATIONAL DEFENSE
RESEARCH INSTITUTE

Best Available Copy

The research described in this report was conducted within two RAND federally funded research and development centers: Project AIR FORCE, sponsored by the U.S. Air Force, Contract F49620-91-C-0003, and the National Defense Research Institute, sponsored by the Office of the Secretary of Defense and the Joint Staff, Contract MDA903-90-C-0004.

ISBN: 0-8330-1383-1

The RAND Publication Series: The Report is the principal publication documenting and transmitting RAND's major research findings and final research results. The RAND Note reports other outputs of sponsored research for general distribution, including RAND Graduate School Dissertations. Publications of RAND do not necessarily reflect the opinions or policies of the sponsors of RAND research.

**Published 1993 by RAND
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
To obtain information about RAND studies or to order documents,
call Distribution Services, (310) 393-0411, extension 6686**

R-4211-AF/OSD

Modeling and Forecasting the Demand for Aircraft Recoverable Spare Parts

John L. Adams, John B. Abell, Karen E. Isaacson

Prepared for the
United States Air Force
Office of the Secretary of Defense

RAND

Accession For	
NTIS	CRA&I <input checked="" type="checkbox"/>
DTIC	TAB <input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification _____	
By _____	
Distribution / _____	
Availability Codes	
Dist	Avail and / or Special
A-1	

Approved for public release; distribution unlimited

PREFACE

This report explores issues in forecasting and modeling the demand for aircraft recoverable spare parts to improve the Air Force's estimation of spares and repair requirements over quarterly, annual, and longer planning horizons. Specifically, it demonstrates the utility of approaches that account explicitly for nonstationarity and their superiority over current methods used by the Air Force Materiel Command for these purposes. It is part of a larger body of research, carried out in the Logistics Requirements Project, and is intended to enhance our understanding of the implications for requirements estimation of demand uncertainty and logistics management adaptations to cope with it. The several reports that describe this work are listed here:

- John B. Abell et al., *Estimating Requirements for Aircraft Recoverable Spares and Depot Repair*, RAND, R-4210-AF, 1993.
- John B. Abell and Frederick W. Finnegan, *Data and Data Processing Issues in the Estimation of Requirements for Aircraft Recoverable Spares and Depot Repair*, RAND, MR-264-AF (forthcoming).
- Donald P. Gaver, Karen E. Isaacson, and John B. Abell, *Estimating Aircraft Recoverable Spares Requirements with Cannibalization of Designated Items*, RAND, R-4213-AF, 1993.
- Karen E. Isaacson and Patricia Boren, *Dyna-METRIC Version 6: An Advanced Capability Assessment Model*, RAND, R-4214-AF, 1993.
- John B. Abell, *Estimating Requirements for Aircraft Recoverable Spares and Depot Repair: Executive Summary*, RAND, R-4215-AF, 1993.

The first of these reports describes the main body of work. The second discusses data and data-processing issues related to estimating aircraft recoverable spares and repair requirements. The third presents a computational algorithm for estimating requirements for aircraft recoverable spares based on the assumption that items can be designated as cannibalizable or not. The fourth describes the capability assessment model used to evaluate the stockage postures that were anticipated to eventuate from purchases of particular mixes of recoverable spares. The fifth report summarizes the entire body of

work and discusses the major findings and recommendations that have emerged from this research.

This work had the joint sponsorship of Headquarters, United States Air Force (AF/LEX); Headquarters, Air Force Logistics Command, now the Air Force Materiel Command (AFMC/XP and AFMC/XRI); and the Director of Maintenance Policy, Office of the Assistant Secretary of Defense for Procurement and Logistics. It was carried out in the Resource Management and System Acquisition Program of Project AIR FORCE, RAND's federally funded research and development center (FFRDC) supported by the U.S. Air Force, and in the Acquisition and Support Policy Program of the National Defense Research Institute, RAND's FFRDC supported by the Office of the Secretary of Defense. It should be of particular interest to those concerned with spares and repair requirements estimation, logistics system design and modeling, and logistics policy analysis. It should also interest other persons concerned with modeling certain stochastic processes.

SUMMARY

This report addresses the problem of estimating Air Force needs for aircraft recoverable spare parts and their depot-level repair. Since the forecasting techniques imbedded in the Air Force Materiel Command's current requirements estimation process were implemented, a great deal has been learned about modeling parts demand processes¹ more effectively. The report explores several issues involved in modeling and forecasting demands for aircraft spare parts along with alternative forecasting methods that can substantially reduce expected predictive error. The research described here is part of a larger body of work intended to help us understand better the effects of uncertainty and management adaptations in shaping the performance of the logistics system in a variety of peacetime and wartime scenarios, and to account explicitly for those effects in spares and repair requirements estimation.

In the last decade, logistics research at RAND has focused on combat logistics support. This interest began with particular attention to the period of transition from peacetime to wartime when activity levels were anticipated to increase sharply, thus dramatically perturbing resource demand processes. That nonstationarity in demand prompted the exploration and development of substantial improvements in logistics research tools and approaches, especially in the Dyna-METRIC series of capability assessment models [50-53, 70]. Continued interest in the problems of wartime logistics support led the late Dr. Gordon B. Crawford to undertake a project to understand better the factors that caused particular recoverable aircraft spare parts to become "problem items." It was known popularly as the "Drivers Project" because of its particular concern with factors that tended to "drive" the performance of the logistics system.

One outgrowth of the Drivers Project was Crawford's observation that items identified as problem items tended to exhibit high variability in their demands. That observation led him to further explore and quantify the magnitude and pervasiveness of variability in the demands for aircraft spare parts. Crawford's findings were published in

¹In the Air Force's Recoverable Consumption Item Requirements System, the system used to compute spares and repair requirements, *demands* are defined as removals of components from their next higher assemblies, excluding components that are declared to be serviceable after subsequent bench check. They also exclude part removals to facilitate other maintenance, etc.

January 1988 [2], but the substance of the work was finished in 1985 and helped motivate the Uncertainty Project, a major research effort at RAND that explored the effectiveness of management adaptations in coping with uncertainty, especially in wartime.² The adaptations included cannibalization; lateral supply; lateral repair; priority repair; assured, responsive intertheater and intratheater transportation; and responsive depot repair. The importance of such adaptations was illuminated by Cohen, Abell, and Lippiatt [20] in their summary of the major evaluations and underlying logic of the Uncertainty Project.

The thrust of the forecasting work described here is a different approach, in a sense, than our recent past work, although it extends a larger body of demand modeling work done at RAND in the 1950s and 1960s, and reopened by Muckstadt, Crawford, and Carrillo [2, 49, 55] in the last decade or so. The methodology that emerged from this particular part of the current project is a potentially useful and important improvement to the current system's approach to demand modeling and forecasting.

This work does not resolve the important problem of estimating wartime demand rates. Although the mix of demand rates resulting from the underlying failure process may not necessarily change dramatically in wartime, many of the events that can occur in wartime are essentially unknowable in advance. Better demand forecasting that incorporates Bayesian updating may be helpful in adjusting to changes in wartime more quickly than the current system would but, clearly, improving our demand forecasting ability cannot by itself help us know the unknowable.

The current system uses an eight-quarter moving average of past demands by line item to estimate item demand rates. This approach assigns as much importance to events in the more distant past as to recent events. Moreover, in forecasting future demands, we currently assign no more uncertainty to events far in the future than we do to events in the short term. Our models assume that parts demand processes have certain characteristics which empirical observations tell us they do not have. For example, in general, demand processes are nonstationary; we assume stationarity. They are not, in general,

²The Uncertainty Project was part of RAND's Resource Management Program. It was formally entitled "Enhancing the Integration and Responsiveness of the Logistics Support System to Meet Peacetime and Wartime Uncertainties" and was sponsored by Headquarters, USAF/LEX, and Headquarters, AFLC/KP.

compound Poisson processes.³ The large variability we observe in these processes is typically not due to batching, or compounding, as is so often assumed in the literature and in our models of demand. Certainly, nonstationarity plays an important role in shaping the variability.

Important characteristics of the statistic used by AFMC to estimate the variance-to-mean ratio (VTMR) of parts demand processes are also discussed here. The variance of the VTMR estimator increases as a function of the coarseness of the partitioning of the observed data, even when the demand process is stationary. If the process is nonstationary, the expected value of the estimator *and* its variance increase with the coarseness of the partitioning and with the demand rate, behavior that is consistent with the association of high values of the VTMR with high values of demand rate. These findings suggest that there are estimation problems associated with AFMC's use of the VTMR estimator. In this report, we present an improved approach to specifying the variance that has more satisfying properties than the current model.

An approach to demand forecasting that seems especially appealing on an intuitive level, and that performs well in empirical evaluations, is weighted regression, a special case of the Kalman filter. It is a logical extension to Bayesian statistics that explicitly accounts for nonstationarity in stochastic processes, assigning greater weight to more recent past demands than earlier ones. Coupled with the improved approach to variance estimation which assigns greater uncertainty to longer planning horizons than to shorter ones, it holds the promise of reducing the cost of spares investments while achieving adequate levels of system performance. For planning horizons 10 to 13 quarters long, the improved techniques reduced forecasting errors on high-demand items by roughly 40 to 50 percent, as shown in Table S.1. (Also see Figures 5.3 and 5.4.)

Table S.1

Percentage Improvement in Root Mean Squared Error and Mean Absolute Deviation of Improved Techniques over Current System

Measure	10-Quarter Horizon	13-Quarter Horizon
RMSE	48	38
MAD	51	45

³A definition of a compound Poisson process may be found in the Glossary.

Moreover, in the case of the requirements computation done with the March 1986 D041 database, about \$76 million in procurements of primary operating stock could have been saved with the improved demand and variance forecasting techniques in place with a modest improvement in system performance, as shown in Table S.2.

The improved techniques also enabled an investment reduction of almost a quarter of a billion dollars (\$239 million) with roughly the same performance.

These results are shown in Table S.3. (The details of these evaluations are explained in Section 7.)

In the applications reported here, the improved demand forecasting method incorporates a normal distribution to approximate quarterly demand. Compared to the current system, it performs especially well on items with a mean of 15 or more demands per quarter (high-demand items), yielding the roughly 40 to 50 percent reduction in RMSE and MAD already mentioned. We do not recommend its use on low-demand items. Its performance on low-demand items was not impressive, perhaps due to failure of its underlying assumptions.

Table S.2
Cost and Performance with Traditional Availability Goals

Management Adaptations	Percentage of Aircraft Unavailable, Peacetime	
	Current System, \$3,709 Million	Improved Methods, \$3,633 Million
No cannibalization	74.9	71.7
Full cannibalization	33.0	32.3
Cannibalization, lateral supply	17.3	16.4
Cannibalization, quick, lateral supply	3.2	3.1

Table S.3
Cost and Performance with Reduced Budgets

Management Adaptations	Percentage of Aircraft Unavailable, Peacetime	
	Current System, \$3,474 Million	Improved Methods, \$3,470 Million
No cannibalization	81.5	76.3
Full cannibalization	34.5	33.1
Cannibalization, lateral supply	19.0	17.2
Cannibalization, quick, lateral supply	3.5	3.6

As the literature suggests, estimating demand rates on low-demand items has always been a troublesome topic, and this work does not resolve the problem. The idea of pooling data across many low-demand items to gain strength from additional observations in estimating demand rates for individual items was suggested years ago in earlier RAND work. The idea seems worth pursuing in future research.

For high-demand items, we recommend using a weighted regression technique for demand forecasting and, for all items, an improved method for specifying the VTMR of the probability distribution used to describe the numbers of assets of each type in resupply (i.e., in the pipeline). The weighted regression technique is a more easily implemented version of the general Kalman filter model, especially for systems as large as the Air Force's recoverable item inventory system. Implementing these techniques should reduce the investment required to achieve a specified level of aircraft availability in peacetime by a substantial amount.

ACKNOWLEDGMENTS

We are indebted to our colleagues at RAND, Frederick W. Finnegan and Patricia K. Dey, for producing the many data files to support this work. We are also indebted to Grace M. Carter, Donald P. Gaver, Kenneth J. Girardini, and Louis W. Miller for their careful reading of an earlier version of the manuscript and their helpful suggestions. Some of the material in Section 4 was drawn from material related to a course in Kalman filtering taught by one of the authors with James S. Hodges of RAND.

Jack Hill, Thomas Kramer, Curtis E. Neumann, and Victor J. Presutti, Jr., of Headquarters, AFMC, provided the necessary data and were very helpful to us in our attempts to understand the many intricacies of the current system.

Colonel Douglas J. Blazer, USAF, formerly of Headquarters, USAF, was also very supportive of this work and deserves special thanks for arranging for our acquisition of models and data from the Logistics Management Institute. We are indebted to Michael Slay, Randall King, T. J. O'Malley, and Virginia Mattern of the Logistics Management Institute for kindly providing those models and data.

CONTENTS

PREFACE	iii
SUMMARY	v
ACKNOWLEDGMENTS	xi
FIGURES	xv
TABLES	xvii
GLOSSARY	xix
DEDICATION	xxv

Section

1. INTRODUCTION	1
2. BACKGROUND	4
The Early Years: 1954-1969	4
The Later Years: 1969-1990	20
Demand Modeling in the Current System	26
3. SOME EXPLORATIONS OF BASE-LEVEL DEMAND DATA	30
Item Characteristics That Tend to Predict High VTMRs	30
The Variance of the Estimator, r_j	34
Nonstationarity in Base-Level Demand Processes	35
Implications of the Base-Level Analyses	36
Wartime Uncertainties	38
4. ALTERNATIVE APPROACHES TO DEMAND FORECASTING	39
Exchangeability Uncertainty	40
Model and Structural Uncertainty	41
Data Uncertainty	41
Parameter or Coefficient Variance	42
Stochastic Variance	42
Accounting for Uncertainties	42
The Kalman Filter	43
Introduction to the Simple Kalman Filter	44
Introduction to the General Kalman Filter Model	51

5.	MORE FLEXIBLE DEMAND AND NRTS FORECASTS	57
	The Parts Sample	58
	Methods and Results	59
	The Computational Algebra	62
	NRTS Forecasting	64
	Conclusions	64
6.	VARIANCE ESTIMATION	67
	Some Observations About the VTMR Estimator	67
	Improved Variance Estimation	73
7.	EVALUATIONS OF THE IMPROVED DEMAND FORECASTING AND VARIANCE SPECIFICATION TECHNIQUES	77
8.	CONCLUDING REMARKS	81
Appendix		
A.	RELATIONSHIP OF THE KALMAN FILTER TO OTHER FORECASTING METHODS	85
B.	THE FUNCTIONAL FORM OF THE MEAN DEMAND PROCESS	88
	REFERENCES	95

FIGURES

2.1.	Observed VTMR as a Function of Expected Number in Resupply	28
2.2.	Enlargement of Lower Left Portion of Figure 2.1	28
4.1.	A Step Function	48
4.2.	An Outlier	49
4.3.	A Ramp	49
4.4.	An Oscillating Process	50
4.5.	The Effect of the Prior on the Forecasts	51
4.6.	State Variance and Measurement Variance Equal	54
4.7.	State Variance Set to Zero	55
5.1.	Root Mean Squared Error of Weighted Demand Rate Forecasters for High-Demand Parts	60
5.2.	Mean Absolute Deviation of Weighted Demand Rate Forecasters	61
5.3.	Root Mean Squared Error of Weighted Regression Forecasters	62
5.4.	Mean Absolute Deviation of Weighted Regression Forecasters	63
5.5.	Root Mean Squared Error of Improved NRTS Forecaster	65
5.6.	Mean Absolute Deviation of Improved NRTS Forecaster	65
6.1.	Effect of the Estimator's Bias Function	69
6.2.	Effect of Data Partitioning on the Standard Deviation of the VTMR Estimator	69
6.3.	Effect of Data Partitioning on the Expected Value of the VTMR Estimator	70
7.1.	Schematic Representation of Evaluation System Software	78

TABLES

S.1.	Percentage Improvement in Root Mean Squared Error and Mean Absolute Deviation of Improved Techniques over Current System	vii
S.2.	Cost and Performance with Traditional Availability Goals	viii
S.3.	Cost and Performance with Reduced Budgets	viii
3.1.	Effects of Partitioning on Observed VTMRs (all 973 parts)	36
3.2.	Effects of Partitioning on Observed VTMRs (918 parts with QPAs less than 5 and action quantities of 1)	36
7.1.	Percentage of Improvement in RMSE and MAD of Improved Techniques over Current System	77
7.2.	Cost and Performance with Traditional Availability Goals	80
7.3.	Cost and Performance with Reduced Budgets	80
B.1.	Results of the Variable Selection Search	92
B.2.	Fraction of Outlier Rejections	92
B.3.	Counts of Outlier Rejections Power Transformation	93
B.4.	Transformations of the Explanatory Variables	93

GLOSSARY

A priori. Without prior information.

AFAO. Authorized force acquisition objective. The total number of assets that the spares requirements computation specifies for the entire inventory system.

AFLC. Air Force Logistics Command, now the Air Force Materiel Command.

AFMC. Air Force Materiel Command, formerly the Air Force Logistics Command.

Aircraft Availability Model (AAM). The system of software imbedded in D041 that is used to compute requirements for safety stock of selected recoverable items.

AOCP. Aircraft out of commission for parts. In earlier years in the Air Force, aircraft were AOCP when they couldn't safely be flown because of lack of parts.

Bayesian. A term deriving from the work of Rev. Thomas Bayes (1701-1761) describing an approach to optimal learning combining new data with prior judgments or old data using the laws of conditional probability.

Beddown. A term denoting the allocation of weapons by type to locations.

Bias. The property of a statistical estimator such that its mathematical expectation differs from the numerical value of the parameter it is used to estimate.

BP15. Budget Program 15, a category of appropriated funds allocated to recoverable aircraft replenishment spares.

Compound Poisson process. A stochastic process in which the numbers of arrivals that occur in disjoint time intervals of equal length are described by the Poisson probability distribution, and the number of events that occur with each arrival is described by a separate, usually different, probability distribution. The number of events that occur with each arrival is called the *compounding random variable*.

Consumable. The property of a part or material such that it is discarded after failure or is consumed in use.

CONUS. The continental United States.

CSIS. The Central Secondary Item Stratification, AFMC's system that is incorporated in D041, along with the Aircraft Availability Model, for computing requirements for recoverable spares.

D028. AFMC's Central Stock Leveling System that allocates stock levels for recoverable items to the bases and the depot.

D041. AFMC's system for computing requirements for aircraft recoverable spares, formally entitled the *Recoverable Consumption Item Requirements System*.

Degenerate probability distribution. A probability distribution of a random variable with only one possible value.

Demand. In the Air Force's Recoverable Consumption Item Requirements System, the system used to compute spares and repair requirements, *demands* are defined as removals of components from their next higher assemblies excluding components that are declared to be serviceable after subsequent bench check. They also exclude part removals to facilitate other maintenance, etc.

Empirical Bayes procedures. Bayesian procedures in which observed data are used to estimate the prior distribution in lieu of subjective judgment.

EOQ. Economic order quantity, the requisition quantity that is determined to be the most cost-effective, usually a function of demand rate, reorder cost, holding cost, interest rate, and unit price. Formulations based on shortage cost are also common.

Exponential smoothing. A procedure for discounting observations more heavily the further they occurred in the past by multiplying each precessive observation by an increasing integral power of a number between zero and one.

Gamma probability distribution. A probability distribution of the continuous type whose density function is given by

$$\frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta}$$

Homogeneous process. A stochastic process whose probability distribution is invariant.

James-Stein estimators. A class of empirical Bayes estimators formed by modifying estimates of parameters underlying observations

of realizations of a stochastic process by pooling them with observations from other, similar populations.

LRU. Line-replaceable unit, a part or assembly that is typically removed directly from an aircraft when undergoing maintenance other than adjustment, calibration, or servicing.

MAD. Mean absolute deviation. The average unsigned difference between a set of estimators and the true values of the parameters being estimated.

Master stock number. The stock number assigned to the preferred item in a set of two or more interchangeable items.

METRIC. Multi-Echelon Technique for Recoverable Item Control, a method for estimating requirements for aircraft recoverable spare parts developed by C. C. Sherbrooke of RAND.

MICAP. A term denoting mission capability effect of a part shortage.

Moving average. The statistic formed by the mean of a fixed number, n , of observations of a stochastic process where the most recent n observations are summed and divided by n . As observations accumulate over time, the latest observation is added to the sequence and the $n + 1$ st observation, counting backward, is discarded.

Negative binomial probability distribution. A probability distribution of the discrete type that may apply to situations in which events occur at random but the variance of the numbers of events in nonoverlapping time intervals of equal length is higher than allowed by the Poisson distribution. Its density function is given by

$$\binom{r + x - 1}{x} (1 - p)^r p^x, x = 0, 1, 2, \dots$$

NRTS. Not repairable this station. The designation is given to a repairable part whose repair is beyond the capability of maintenance at a particular location.

Partitioning. The subdivision of a set into subsets that are mutually exclusive and collectively exhaustive.

Poisson process. The most widely known and often used form of stochastic model with important mathematical properties that make it especially tractable and useful. It is described by the Poisson probability distribution whose density function is given by

$$\frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, 2, \dots$$

POM. Program Objectives Memorandum, the programming document used by the military departments to state their budgetary requirements in future fiscal years.

POS. Primary operating stock, formerly known as *peacetime* operating stock.

Power function. A mathematical function of the form $y = ax^b$.

QPA. Quantity per application, the number of parts of a particular type installed on the part's next higher assembly.

Random walk. A stochastic process whose location parameter varies from one realization to the next in a manner determined by chance.

REALM. The Readiness/Execution Availability Logistics Module, the software module of WSMIS that computes requirements for war readiness spares.

Recoverable parts. A class of parts that are repaired when they fail, rather than being discarded or consumed in use.

Resupply. The state of parts that are in base repair, depot repair, shipment from one location to another, or have been condemned and whose replacement is pending.

RMSE. Root mean squared error, computed as the square root of the average squared difference between a set of estimators and the true values of the parameters being estimated, a popular measure of predictive accuracy. Also referred to as RMSD, root mean squared deviation.

Safety stock. Spares authorized to accommodate the variability in the numbers of items in resupply.

Skewness. The characteristic of a probability distribution such that it is asymmetrical about the mean.

Spectral analysis. A method for finding and quantifying periodicity in data.

SRU. Shop-replaceable unit, a subassembly of an LRU that is typically replaced during repair of the LRU.

Stationary process. A stochastic process whose parameters are invariant.

Stock level. Serviceables on hand plus due-ins minus due-outs.

VSL. Variable safety level, a spares requirements estimation method derived directly from METRIC, a method developed by C. C. Sherbrooke of RAND.

VTMR. Variance-to-mean ratio, defined as the unbiased estimator of the variance of a process divided by its mean.

WRSK. War readiness spares kit, an airlift-deployable set of spares to support squadrons deployed in contingencies.

WSMIS. AFMC's Weapon System Management Information System.

DEDICATION

We respectfully dedicate this work to the memory of our RAND colleague, Dr. Gordon B. Crawford, who perished in an aircraft accident in December 1989. Gordon first suggested this research to one of the authors and proposed the original organization of this report. He also carried out the initial formulation and evaluation of the idea of using the Kalman filter on spares demands. His early work with a few high-demand items was compelling in its implications for further exploration of the Kalman filter. The work described in this report merely extends and enhances Gordon's original thinking.

Gordon had long been interested in understanding better the mechanisms underlying the demands for aircraft spare parts. He had also conducted some research in the past on the use of James-Stein estimators for adjusting base-specific demand rates with worldwide means and had explored the use of alternative models of the relationship between demands and flying hours. His interests also extended to the problem of estimating wartime demand rates, which, although not specifically addressed in this research, remains an important unresolved issue.

His intellectual mark is clearly imprinted on this work.

1. INTRODUCTION

This report addresses the difficult problem of forecasting Air Force requirements for aircraft spare parts and their depot-level repair. These forecasts are made over different time horizons according to the specific purpose of the forecast. For example, the Air Force is required to forecast its budgetary requirements for these resources across the multi-year horizons involved in the POM. AFMC also forecasts annual spares requirements and annual and quarterly repair requirements. The difficulty of making such forecasts has two fundamental roots: (a) substantial variability in spares demands, even in peacetime (statistical uncertainty), and (b) instability in force structure, force beddown, flying hour programs, funding profiles, item reliabilities, and other item characteristics (state-of-the-world uncertainty).

The levels of variability in peacetime demands will probably be compounded in wartime by system disruptions, resource losses, and the inevitable surprises of combat, exacerbating the demand forecasting problem. (The estimation of wartime demand rates is not addressed in this work.) The robustness of the spares postures delivered by alternative approaches to spares and repair requirements estimation, a central issue of the larger body of research of which this is one part, is reported elsewhere, as mentioned in the Preface. RAND has devoted considerable attention in recent years to the role of management adaptations, e.g., cannibalization, lateral supply, and priority repair, in overcoming uncertainties in resource demands in peacetime and wartime. That research and its relationship to this work are discussed briefly in Section 2.

It is difficult to make accurate forecasts of spares and repair requirements. On more than one occasion, the Air Force has had to adjust its budgetary requirements for these resources as the execution year approached, inducing considerable turmoil in the resource allocation process. Lippiatt noted that current requirements and capability assessment systems do not explicitly consider parameter variabilities and forecasting uncertainties [1, p. vi]. The forecasting algorithm of AFMC's spares and repair requirements estimation system can be improved. The current system uses eight-quarter moving averages to estimate item demand parameters, a technique that gives no more weight to relatively recent observations than to older, often less relevant observations. The estimates that emerge from this approach do

not reflect the nonstationarity commonly observed in these demand processes. Moreover, the current system assumes a strictly linear relationship between demands and flying hours. Thus, if the planned flying hour program increases by 10 percent, the system's estimates of expected demands also increase 10 percent. Unpublished analysis suggests that the linearity assumption is not only incorrect, under some conditions it is grossly incorrect in attributing too large an effect to flying hour changes.

Crawford characterized the magnitude and pervasiveness of variability in peacetime demand [2]. He noted that, in estimating war readiness spares requirements and in capability assessment modeling, levels of variability in demand were typically understated. In unpublished research, he also noted that observed demands were not linearly related to flying hours; he estimated a much smaller effect of flying hours on demands.¹ Thus the assumptions of the current system are in serious doubt. Crawford hypothesized that the use of the Kalman filter, coupled with an improved model of the relationship between demands and flying hours, might be a more appropriate approach to modeling demands for aircraft spare parts. The research reported here examines and confirms Crawford's hypothesis about the Kalman filter and recommends the use of a special case of the Kalman filter, weighted regression forecasting, for estimating spares demands. The techniques developed here can be expected to achieve specified levels of system performance at less cost through more reliable demand forecasting.

In Section 2 we discuss, by way of background, the larger context of this research, its relationship to past RAND research, and AFMC's current approach to modeling demands in its estimation of spares and repair requirements. We describe some base-level demand experience in Section 3 and discuss its implications for demand modeling, spares and repair requirements estimation, and inventory management. In Section 4 we provide an elementary description of the Kalman filter and its historical derivation and describe our formulation and evaluation of alternative approaches to demand forecasting. We describe our evaluations of some alternative demand forecasting techniques in Section 5, including weighted regression forecasters, which are a special, limiting case of Kalman filter regression forecasters. We discuss an alternative model of variance specification in Section 6 and

¹In a conversation with one of the authors shortly before his death, Crawford reported exploring a postulated model of the form $\ln D = a + b \ln H$, where D is demands and H flying hours. His estimate of the value of the coefficient b was 0.2. To the best of our knowledge, the estimate was made with F-16 data.

present our evaluations of an improved demand forecasting method and an alternative specification of variance in Section 7 using a capability assessment model and replicas of the Air Force's spares requirements computation and asset allocation systems. We offer some brief concluding remarks and recommendations in Section 8.

2. BACKGROUND

For several years beginning in the mid-1950s, RAND researchers pursued the problem of forecasting demands for aircraft spare parts. That interest was sustained through 1969 when George Fishman published a Research Memorandum (RM) on improved forecasting methods and thus ended for more than a decade the publication of any research on demand forecasting. The next publication in the RAND literature on the topic appeared in 1980. It addressed the issue of nonstationarity in spare parts demand processes, a topic that absorbed considerable attention among RAND's logistics researchers throughout the 1980s. Two recent publications have, in an important sense, synthesized the principal thrusts of the thinking of RAND researchers over these four decades. We will summarize them in due course. It is interesting to note that this current work has roots in the earlier RAND research of the 1950s and 1960s as well as that of the 1980s. In the pages that follow, we summarize the work at RAND with special emphasis on those issues that relate most directly to our current thinking about this important topic.

THE EARLY YEARS: 1954-1969

RAND's earliest work in demand forecasting was published in July 1954. It was made possible by a special data-collection effort that provided RAND researchers with about 1,300 aircraft months of data describing demands for spare parts at three B-47 bases. That particular database absorbed the attention of several researchers for several months. It contained spares demands generated by about 230 aircraft and 33,000 flying hours. Before the special data-collection project that generated these data, data describing demands by "aircraft model" (weapon system) had been unavailable. Only data describing *issues* of spares had previously been available but not by weapon system, only by "property class" (federal stock class) that overlapped weapon system. Issues, of course, did not reflect *backorders*; thus, in these early years, analysts were somewhat constrained in their ability to identify and characterize the nature of true demands. Geisler, Brown, and Hixon [3] summarized the earliest analytic findings based on the B-47 data as follows:

It is shown that there was a surprisingly low amount of demand both as to kinds and quantity of aircraft spare items, at either MacDill, March or Fairford (England) Air Bases, compared with the number of kinds of

such items in the United States Air Force (USAF) Supply Catalogue, and the quantities of such items stocked. Furthermore, we could find no significant correlation between the kinds or quantity of items demanded and the flying activity of the aircraft, measured in either flying hours, landings or aircraft months. We also found that a comparatively small per cent of the reparable turn-ins of airframe and engine spare items were repaired at the one base studied (MacDill Air Force Base) with most of the items either condemned at the base or sent to the depot for repair.

These conclusions are hardly surprising in retrospect because the B-47 database included both consumables and recoverables. It is important to note, however, the lack of "significant correlation" between demands and activity levels. Shortly after this first publication, Brown and Geisler [4] reported their analysis of demand data for B-47 airframe parts:

Both the daily demand for individual kinds of items and the daily combined demand over all items show more variation than expected from the Poisson distribution, which was used as the theoretical model of demand. . . . These results indicate that if the Poisson distribution is used to represent the demand pattern for spare items because of its mathematical convenience, the actual distribution for either individual items or combined may be more extreme, in that the variance of the distribution will be greater than the mean value of demand.

Thus, it was recognized from the start that demands for aircraft spares exhibited unexpectedly high variation. Again, the authors failed to differentiate between consumables and recoverables; however, they did observe differences in demand patterns between high- and low-cost items.

Geisler and Youngs [5] explored one implication of the B-47 demand data analyses for base supply stockage policy. They suggested that expensive items with low demand rates should not be stocked at the base. Thus the use of cost as well as demand as a criterion for stockage also emerged right from the start. Geisler and Youngs made other fundamentally important observations in this early paper. They discuss at some length the problem that past demand may not shed much light on an item's "true" demand rate. They say, "In point of fact, one of the most troublesome problems is that of obtaining some grip on the *true demand rate*. This is particularly true of the host of items (10,000 strong) in the O1A-FE category which have never been demanded" [emphasis added]. The idea that items have nonstationary demand rates does not emerge until considerably later. After exploring the implications of the decision to stock an item with low or no observed demand in some past period, the authors point out

that "as a matter of general policy it may be argued that the decision on whether these expensive items are to be stocked should not be undertaken at base level where the data are not extensive (and therefore subject to more variations) in comparison with the data at depot level. This question is much affected by the stockage policy assumed by the depot when it determines the worldwide requirements and establishes procurement needs." This observation seems consistent with the Air Force's later implementation of central stock leveling. Ironically, however, the central stock leveling system uses base-specific data reflecting past demands over a specified period of observation.

A short time later, Brown and Geisler [6] examined data describing spares demands on B-50D aircraft from a wing deployed to RAF Station Upper Heyford, England, during a 90-day period, and contrasted them with demands on B-47 aircraft deployed to Fairford, England, for a similar period. Little of importance was noted beyond the larger numbers of demands per flying hour experienced by the B-50s, especially for small hardware items, not surprising in view of the difference in technologies represented by the two types of aircraft. Karr [7] related the B-47 database describing spares demands with data describing AOCP (aircraft out of commission for parts) occurrences. He analyzed eight weeks of AOCP reports from two B-47 bases (March and MacDill). AOCPs are roughly equivalent to MICAP occurrences in the modern Air Force. His observations seem consistent with the distributions of parts shortages among aircraft to this day, and serve to point out the important role of consolidating shortages (cannibalizing) among aircraft as well as the inhibiting effect of shortages of parts that cannot readily be cannibalized. His conclusions included the following observations:

1. Most of the shortages were corrected within a few days (91 per cent appeared in only one weekly report).
2. Most of the out of commission aircraft lacked only a single part during the eight week period studied (60 per cent lacked one part, 23 per cent lacked two parts).
3. There were several aircraft in a hangar queen condition (11 lacked 5 or more parts).
4. A large proportion of the aircraft were out of commission because of lack of parts during the eight week period studied (71 per cent of all aircraft on the two bases studied were out of commission at one time or another).

5. The demand rates for the lacking parts were typically very low (52 per cent had experienced zero demand during a 1300 aircraft-month sample).
6. Most of the lacking parts were very inexpensive (37 per cent cost less than \$10).

Our guess is that, with an adjustment for inflation, Karr's observations would apply quite well to the current world.

With the realization that demands for aircraft spares were not adequately described by the Poisson distribution, Youngs, Geisler, and Mirkovich [8] published an RM describing confidence intervals for Poisson parameters in logistics research. On the face of it, this would seem to ignore the obvious, but it was simply a first step in a move *away* from the Poisson to the negative binomial probability distribution. It was, in fact, an explicit recognition of the fact that past demands may not accurately represent an item's demand rate.

The first explicit reference in the early RAND literature to the use of negative binomial probability distributions to describe spares demands occurred in an RM by Youngs, Geisler, and Brown [9]. In this case, the negative binomial distribution was used to describe the conditional probability of observing y demands in some specified future time period having observed x demands in some past period. The experience base was chosen to be aircraft-months of experience. The negative binomial model was based on the assumption of a Poisson demand process with a gamma-distributed location parameter, a result published in 1920 by Greenwood and Yule [10]. In their opening discussion, Youngs, Geisler, and Brown conclude that:

The method of conditional probabilities is much more precise than the straight-forward Poisson approach for the low probabilities and frequencies of demand, which is very important in the case of aircraft spares. For the higher demand rates, (demand rates greater than 0.5 per 100 aircraft months) the two approaches give very similar results. Also, the results of the two methods converge as the experience period gets longer.

The conclusion about convergence was based on the assumption of a steady-state process involving an unknown constant location parameter. It was not based on the data. Given what we have subsequently observed about spares demand processes using large datasets, it is incorrect.

In February 1955, Geisler [11] documented a briefing he gave to the Long Range Logistics Research Conference in Santa Monica. He

summarized some of the analyses of demand data RAND had done up until that time and suggested the desirability of increasing the stock levels of low-cost items at base level. Shortly after Geisler's briefing, he and Mirkovich [12] published analyses of spare parts demands on F-86D aircraft and noted many similarities with those on B-47 aircraft. The same authors [13] reported analyses of worldwide spares demand data that reinforced the observations made by Geisler in his briefing. A short time later, Hamburger [14] suggested the evaluation of simplified distribution functions to describe item demand. His work apparently received little attention; however, in comparing the simpler functions that he suggested with the Poisson distribution, he wrote:

Demand predictions are affected by three major sources of error:

- 1) The random occurrence of demands: Actual demands fluctuate about the true demand rate in a fashion described by the Poisson distribution. These fluctuations are particularly noticeable when periods of time which are short, relative to the demand rate, are considered (e.g., the number of automobile accidents occurring over a weekend can be predicted more precisely than the number occurring in the course of an hour).
- 2) Insufficient information for determining true demand rates: It may be that the true demand rate of parts with an experienced demand of 5 is actually 10, on the average. But some of these parts will have higher than average true demand rates, while others have lower ones.
- 3) Biased estimates of true demand rates: The estimated average true demand rate of a group of parts may be incorrect.

It was not the first allusion to the idea of a "true demand rate," nor would it be the last.

An RM by Karr, Geisler, and Brown [15] was the first of several that explored the design of flyaway kits (war reserve spares kits in the modern Air Force). Hamburger [16] also explored an alternative approach to computing flyaway kits. Clark [17] offered an approach to the central allocation of stock levels among bases and the depot in the anticipated future Air Force environment in which computers would be available to carry out the tedious computations involved. It is interesting to note that Clark's approach drew on the use of worldwide data to support such allocations consistent with the earlier observation by Geisler and Youngs that base-specific demands were too sparse for such purposes.

Berman [18] postulated a linear programming approach to modeling the procurement-repair decision for an item that tried to account for its life-cycle costs.

The notion that low demand implies unpredictability pervades this early literature. Time and again sparse demands were blamed for the inability to forecast future demands and estimate the demand distribution adequately. In later years, advances in probability theory, specifically Bayesian formulations, brought to researchers a more constructive view of the problem.

The sequence of RMs described thus far culminated in the publication of a formal RAND report by Bernice Brown in July 1956 [19]. Entitled *Characteristics of Demand for Aircraft Spare Parts*, it summarized the work that had been done until that time. The Summary of Brown's report is of special interest here. Excerpts follow.

Knowledge of demand for aircraft parts is needed for effective and economical procurement, distribution, and stockage decisions. The following paragraphs summarize the results of RAND research on demand and point to certain conclusions that can be drawn for the logistics system.

Low average demand rates are characteristic of a large proportion of all aircraft parts. During a year's period at the bases studied (and perhaps at all bases), as many as one-third of the available spare parts had no demand, and three-fourths had so few demands that they offer an unreliable basis for predicting future demand. Moreover, many of the parts had low unit costs. Forty-two per cent of all line items in the *USAF Worldwide Stock Balance and Consumption Report* for 1952-1953 had fewer than ten issues during the year and cost less than \$10 each. The slow-moving, low-cost parts account for a small fraction of the total dollar value of issues, but because of their large number and, often, their essentiality to the functioning of the aircraft, they constitute a significant logistics problem.

Demand for most spare parts also tends to be erratic. Even if the demand rate for a part is known for some past period, the future demand during a similar period cannot be predicted with accuracy. To reduce the occurrence of parts shortages to a reasonably low level, it is not enough to predict (and use) average demand rates, but, rather, it is necessary to predict the probability that various demands will occur. These probabilities can then be used in calculations designed to facilitate stockage, procurement, and other important logistics decisions.

In many cases, the pronounced random element causing uncertainty in demand can be expressed by a mathematical formula. Many of the airframe parts—with sufficiently frequent demands to permit statistical analyses—show demand patterns that can be approximated by standard probability distributions, such as the Poisson probability distribution. When such approximation is possible, logistics decisions can be

computed that take into account the costs of incurring shortages as well as the costs of avoiding them. This can result in stockage, procurement, and other logistics decisions that produce greater combat capability for the resources available.

There are many causes of the unpredictability of demand for spare parts. . . .

The fact remains that demand for most spare parts cannot now be predicted with confidence, and perhaps never can. This makes it necessary to consider some improvements in logistics operations to make it easier to live with demand uncertainty. Among such improvements would be a shortening of the resupply time, of the procurement lead time, and of the repair cycle for spare parts. Each of these improvements would help to reduce the time over which predictions must be made and would lessen requirements for procurement, thus reducing the risk attending prediction:

1. Shortening resupply time would generally reduce the amount of buffer stocks that must be kept at air bases. These buffer stocks are now large and costly because it is so difficult to predict demand at base level. Even though it might cost the system more to reduce resupply time, the savings in required buffer stocks, as well as the reduction in lost performance time for aircraft suffering the shortage, might outweigh this increase in cost.
2. Reducing the procurement lead time would promise considerable economies in the procurement of spare parts. In the early stages of production, when there is little demand experience and much statistical uncertainty, short procurement lead times, with the option of frequent reorder, would help much to economize on procurement. It is also obvious that demand prediction at all stages is hampered by many dynamic elements, such as unexpected changes in aircraft configuration, in engineering design, and in aircraft procurement schedules. A shortening of procurement lead time would reduce the impact of these elements. Such shortening might be hard to achieve because it would probably require changes in contractual and procurement techniques used by the U.S. Air Force. RAND is doing research along these lines. Very likely, reductions in procurement lead times would be accompanied by increases in unit cost, but these increases should be more than counterbalanced by reductions in the volume of parts procured.
3. A shortening of repair-cycle time, finally, could probably be accomplished only by major revisions in the present system of scheduling and doing repair. This shortening would have the same benefits as the shortening of procurement lead time. In the early stages of production, fast repair would permit the system to operate with a smaller inventory of parts; and at all stages, it would cushion the uncertainties of demand. The ability to repair quickly would require much more rapid transmission of data between bases and depots and more immediate reactions by the depots to such data. Such revisions in the repair system would undoubtedly result in higher unit repair costs, but here again the

final outcome would probably be a net reduction in total cost to the system through less procurement and fewer shortages.

These improvements in logistical management, if made, would probably have valuable effects in various directions that lie outside the scope of this study. In relation to the forecasting of demand, they would tend to overcome some of the costly effects of the very limited predictability of demand for aircraft spare parts.

These remarks have a remarkable resonance for those of us involved in military logistics research at RAND today, 36 years later. We have essentially come full circle, having recently rediscovered, in a sense, the core issues involved in effective logistics support of military operations. In recent years, our motivation has been to identify and evaluate alternative approaches to improving the logistics system's ability to cope with uncertainty. In some ways, our current view of the problem is unchanged from that which emerged from the work of Brown, Geisler, and others in the mid-1950s; however, the source of uncertainty is different, at least in part. It derives not only from the difficulty in predicting peacetime demands, but also from the uncertainties of wartime. As was pointed out by Cohen, Abell, and Lippiatt [20], the unpredictability of resource demands in peacetime is likely to be compounded in wartime by system disruptions, resource losses, and the inevitable surprises of the combat scenario. The solution directions, however, are essentially the same as those suggested by Brown: flexibility and responsiveness in the logistics system to help mitigate our inability to forecast resource demands. This is the principal thrust of our work in spares and repair requirements estimation. Nevertheless, we have explored, and continue to explore, our ability to improve the modeling of peacetime demands for aircraft spare parts, a problem to which we are bringing a collection of methods and ideas that were not well known at the time of the earlier RAND research. Thus, as remarkably insightful as the Brown Summary is, much has been learned in the logistics research community in the intervening years that now enables us to add to the body of knowledge that emerged from this early work, although the principal thrusts of the conclusions in the Brown Summary seem to be as appropriate today as they were then.

Remarks found throughout the early publications reflect implicit assumptions about the uncertainty in demands for aircraft spare parts that, although important to any effort to understand them, are not made explicitly clear. Those assumptions are:

- Items have true demand rates which are unknown constants. If we had enough data, we could estimate them better.
- Sound decisions about stockage of items with no or low observed demand are intractable because we are unable to estimate probability distributions to describe their future demands; therefore, we are likely to make relatively larger errors in predicting demands for these items.
- Items with large numbers of demands in past time periods are more tractable because we are able to estimate the probability distributions of their future demands, but we still have to contend with the problem that we don't really know their true demand rates.

As we will show in the sections that follow, each of these several fundamental assumptions is incorrect. There is a simple observation to make about the latter two of these assumptions. It has to do with relative error as opposed to absolute error. If I predict one event and two occur, I will be wrong by 100 percent. If I predict 100 events and 150 occur, I will be wrong by only 50 percent. The outcome in the second case hardly makes me a better forecaster. If every event that contributes to my forecasting error costs me \$1,000, I would hate to be faced with the latter outcome. Similarly, if we have larger relative error in forecasting demands for low-demand items, it won't necessarily have serious adverse effect on the performance of the inventory system. In fact, fairly modest relative errors in forecasting demands for high-demand, high-cost items can be far more costly either in performance degradation or unwise investments.

An important motivating factor for the interest in low-demand items may have derived from the study by Karr [7] in which he analyzed aircraft AOCF data from March and MacDill AFBs from 9 April to 4 June 1954. The B-47 demand data described above were also used in this analysis. That database covered a different time period at each of three bases; each time period ended in 1953. He found that 50 percent of the AOCF occurrences in the eight-week period he examined were for parts that had no observed demand in the 1,300 aircraft-month database. An additional 11 percent had only one demand in the earlier data. Karr concluded that, "The reason why these parts predominate in spite of their low demand rates is that there are so many of them." Indeed, the distribution of worldwide demands over stock numbers today exhibits similar characteristics, at least for recoverable items.

Another observation made in the early days deserves explicit mention here: "We could find no significant correlation between the kinds or quantity of items demanded and the flying activity of the aircraft, measured either in flying hours, landings or aircraft months." [3, p. ii] In fact, the work done before Brown's report used aircraft months as the denominator of the demand rate.

There are several important observations to be made about this early work:

- All of the analyses published, with the single exception of that of issue data from the USAF Worldwide Stock Balance and Consumption Report by Geisler and Mirkovich [13], focused on base-level demand data rather than systemwide data.
- All of the published work, without exception, pooled data on consumables and recoverables and did not explicitly comment on differences between them.
- No explicit consideration was given to interchangeability and substitutability relationships among line items in the inventory system.

Later work by Goldman [21] pointed out several errors in the approaches taken until that time. His three principal conclusions, although obvious to any researcher in the field now, illuminated the problem considerably at the time:

1. The family of parts rather than the individual part number should be the basic unit in demand analysis and forecasting.

Use of the family of parts, consisting of the master part number and all subsidiary part numbers, as the basic unit of analysis makes it possible to take substitution relationships among line items into account in interpreting consumption data. . . .

2. When the data are analyzed in accordance with the foregoing considerations, future demands for spare parts can be predicted from program-element data.

Goldman [22] subsequently conducted an experiment involving the traditional initial provisioning problem of predicting demand rates for parts without previous consumption experience. He later extended his work [23] to a priori demand prediction for F-100 airframe parts during the acquisition phase of that program.

The problem of initial provisioning and support of weapon systems in the early stages of their life cycles was a topic that absorbed consider-

able attention from RAND researchers in the early years. Some of the inferences about parts demand in the early work found their way into the failure model described by McGlothlin et al. [24], and used in Laboratory Project I (LP I). LP I was a simulation study of alternative logistics support policies and strategies whose simulated performance was to be compared with the support policies and strategies traditionally employed in the acquisition phase of a weapon system's life. Logistics support in the weapon system acquisition phase provided much of the context for this early work; thus the concern of these researchers was not simply with the statistical uncertainty associated with the world of replenishment spares estimation but with a world in which data are sparse and the difficulty of the estimation problem thereby compounded.

The initial provisioning problem provided the context for an RM by McGlothlin and Radner [25] which suggested the use of Bayesian techniques for pooling early observations of demand with initial estimates of demand rates in a systematic way to give the proper weight to observed data in revising the initial estimates in the early life of a weapon system. In a subsequent RM, McGlothlin and Bean [26] suggested procedures for implementing the Bayesian approach.

An RM by Astrachan, Brown, and Houghten [27] reported mixed results from the application of seven different predictive techniques to Falcon missile and B-52 parts. No particular technique emerged as clearly superior to the others.

On 25 and 26 January 1962, RAND sponsored a Demand Prediction Conference which was held at Stanford University. It was attended by several distinguished academicians and researchers who were or had been involved with problems in demand prediction. Kenneth J. Arrow contributed a paper after the conference which was published in an RM edited by Astrachan and Cahn [28]. Arrow's paper summarized the discussions that took place during the conference and included some observations of his own. His observations seem to be especially pertinent and incisive in the context of the present work. Arrow commented:

It is not easy to form an *a priori* opinion about the fruitfulness of statistical forecasting techniques. This needs to be done empirically. In one way or another, most of these methods, apart from spectral analysis, seem to be of the discounted least-squares type.

For what model of the world would this statistical method be correct in any sense? When we consider that we are getting a set of observations from virtually the same universe, why is every observation not as good

as any other one? It is not clear why one should weight them in any way.

The interpretation made by Winters was that we can think of the parameters as shifting, but shifting in a random-walk manner, which adds one more unknown. Looking ahead, things are getting more and more uncertain, and discounting compensates for this growing unreliability. By the same token, if we start from the present the past data are more uncertain. Thus, if we arrived where we are now by a random-walk process, we can also go backwards by the same process. This may be the rationalization for this kind of least-squares method. We have a past marked by change; we think change will persist as we go further into the future. The discounted least-squares methods compensate for this process in some way.

Brown's arguments have shown the flexibility of this model. One can build a great deal into it, apparently, much more than by straight exponential smoothing. Furthermore, it is possible to bring in any explanatory variables we like, such as program elements and age of parts. A combination of smoothing techniques may produce better results with different program elements.

Another point raised concerned the program elements themselves. Assuming that usage does have established relationships to some program elements, then in order to forecast usage we also have to forecast the program elements. Doing so introduces additional "noise." Brown's argument is that it is better to use, as explanatory variables, mathematical functions of time about whose extrapolation there is no question.

There is a counterargument which depends on the use you can make of the forecast. A conditional forecast gives some information that an unconditional forecast does not. It tells us what will happen if we change our minds, so that we might say, after looking at it, "It's really too expensive to fly those things around. I'd better not do it." This would make our forecasting worse, but it would also answer a question that could not be answered with an unconditional forecast. . . .

Arrow's comments, like Goldman's earlier observations, seem obvious, but only in retrospect. What we know the most about is what is happening now. Data from the distant past may be less pertinent than data we have from the recent past; similarly, the longer our planning horizon, the less reliable our sense of the future, and the broader our confidence intervals need to be. Moreover, we may discover an unconditional forecasting method (i.e., one that is not related to program values such as flying hours) whose performance dominates all others by some error measure, but if the world is changing, we are more likely to make sensible forecasts with a conditional method even at the expense of larger expected errors.

Campbell [29] examined demand data from the Air Force's maintenance data-collection system for a squadron of B-52 aircraft over a four-month period using multiple correlation and regression analysis to explore relationships between demands and seven operational variables: sorties flown, flying hours, flying hours at low altitude, bombing-navigation training units, fire control system usage, ECM system usage, and periodic inspections. He concluded that demands seemed to be related to flying hours and sorties, with flying hours having the stronger relationship. Campbell's sample included recoverable items only, not consumables, apparently the first such distinction made in the earlier RAND research. He made an important observation in his closing remarks:

The sharply declining predictability of component demands at lower levels of aggregation suggests a concluding thought. Some information is contained in the prediction by major system aggregate or shop aggregate that is lost when only separate predictions are made. Aggregate predictions can never replace line-item predictions for all support functions, but we must find wider applications for them.

Feeney, Petersen, and Sherbrooke [30] described the evaluation, using actual base demand data for recoverable items observed at Andrews AFB, of a base stockage policy that incorporated demand rates estimated with Bayesian procedures that seemed to respond directly to Campbell's ideas for aggregating data across items. Stock levels were computed using six months of past demands and were then evaluated using demands observed during the subsequent six months. The aggregate fill rate of the computed stock levels was higher than that of both authorized and on-hand stock levels, and differed from predicted performance by less than 5 percent, a difference that declined with postulated investment level. It was a landmark study in the sense that it combined the results of demand prediction work done up to that time with a Bayesian approach to the determination of stock levels that pooled information across line items, a procedure that had previously been done using a line-item-by-line-item approach. They also pointed out the sensitivity of system performance to variability in item demand. The authors made an important observation about demand variability in their closing remarks:

Because any stockage policy must operate implicitly or explicitly with some assumption of demand variability, it is important to note how sensitive stock requirements are in this area. Unfortunately, there is a lack of data with which to estimate base demand variability, and more important, we have little understanding of what causes such extreme fluctuations in demand. Perhaps a large part of variability is simply erroneous reporting. If so, improved reporting quality could produce large

reductions in stock investment. Clearly, we need to know more about the nature and causes of fluctuations in demand. Without such information, we are not in a position to decide what portion of demand variability should legitimately be covered by base stockage policy. It is hardly necessary to point out that this is a basic question, which from a management point of view is as important as choosing a specific target base fill rate.

As noted in previous studies, more responsive resupply does reduce the amount of stock required to achieve a given fill rate. From this it follows that in instances where there is a significant difference between base repair cycle length and depot resupply time, it would be beneficial to establish stock levels for an item as a function of percentage issues base repaired.

No stockage policy can eliminate stockouts; expedited deliveries will still be required from time to time due to the vagaries of demand. But the achievement of a specified target base fill rate with minimum stock investment will cause the costs of expediting to be incurred in resupplying the high-unit-cost items, which represent a more efficient use of support resources.

As we will see, this RM was a harbinger of the important body of work that would ultimately emerge from the collaboration of Feeney and Sherbrooke.

McGlothlin [31] refined and simplified his earlier Bayesian approach to weighting initial estimates of demand with observed demand early in the operational life of a weapon system.

Fishman [32] described the application of spectral analysis to base-level demand data and its usefulness in separating trends and cyclical effects from random events but the method never found its way into Air Force use. Astrachan and Sherbrooke [33] evaluated the use of exponential smoothing in forecasting demand and concluded that it "... does not appear to be a significantly better predictor than the cumulative issue rate techniques currently being used." Their assessment was based on datasets with particular characteristics that may have contributed to this outcome. As we will show in this report, the simplest version of the Kalman filter is essentially exponential smoothing, and we have found it very effective in reducing the mean absolute deviation in demand forecasting. The conclusion reached by Astrachan and Sherbrooke may also have been due in part to the lengths of the time periods involved and the choices of weighting factors.

Feeney and Sherbrooke [34] described the application of Bayesian inference to the analysis of spare parts demand using a different approach from that of McGlothlin. Where McGlothlin was inferring an

estimate of the demand rate assigning appropriate weights to initial estimates and observed data for a single item, Feeney and Sherbrooke suggested an approach that estimated an item's demand rate by observing the distributions of demand rates of all items. They argued that substantial improvements could be made by using system-oriented approaches to supply management and demand modeling over traditional item-oriented approaches and demonstrated elementary examples of this thinking. The approach they described has fundamentally important implications for the demand modeling problem. They make the following observations in their summary:

The traditional approach to demand analysis calculates an item's *issue rate* (demand observed over some past period divided by the length of the period) and assumes that future demand will be some random variation around that level. Such an approach is adequate if the item has relatively high demand, if a relatively long history is available, and if the history is relevant. But many of the most important items, particularly high-cost, low-demand spare parts, fail to meet one or more of these three essential requirements.

A new approach to demand analysis, based on a mathematical technique called *Bayesian inference*, is described in this Memorandum. This approach exploits the surprising fact that we can increase our knowledge of an item by analyzing the behavior of the other items in the system. Instead of trying to estimate the item's true average demand, this approach estimates the probability that the item's average demand is at one of several levels. We think this is a much more realistic way of characterizing what we can learn from an item's demand history. These probabilities can now be used to appraise the true risks and potential payoff for various stock levels. Data-processing procedures can be designed to use information of this kind with little or no increase in processing complexity.

Bayesian demand analysis has two immediate implications for improved supply management. First, because it wrings maximum relevant information from available demand data, this approach promises substantial improvement in the efficiency of stockage decisions. Specifically, by using this kind of analysis we should be less liable to overestimate demand and buy too much because of a random surge in demand, and less liable to underestimate demand and buy too little because of a random decline in demand. Second, because the approach can be applied to any time period, it is extremely flexible. It should be possible to eliminate many of the policy problems now created by items that do not meet the requirements of daily-issue-rate computation: items with low demand, and items with erratic demand patterns. In the framework of Bayesian demand analysis, policies can be developed that prescribe action unambiguously.

Feeney and Sherbrooke [35] extended the well known queuing result of Palm [36] to the problem of determining stock levels for recoverable

spares under the assumption of compound Poisson demand, a much richer case than was previously tractable, although not one on especially solid logical ground, as we shall see later.¹ Feeney and Sherbrooke [37] further discussed the application of Bayesian inference to the demand estimation problem and suggested a method for coupling it with exponential smoothing where the demand rate is nonstationary. They subsequently discussed an approach [38] to setting base stock levels for recoverable spares that integrated the work previously described in Refs. 35 and 37.

Press [39] developed empirical Bayes estimators for the univariate and multivariate exponential family of distributions, for distributions with nuisance parameters, and for the distribution of a family of random variables (Poisson process), extending the work of Robbins [40], and discussed their application to demand modeling. Campbell, Lu, and Michels [41] extended the approach of Feeney and Sherbrooke to the computation of war reserve spares requirements in the context of the aircraft dispersal policy then in effect in the Air Defense Command. Houghten [42] explored alternative procedures for ensuring that AFLC's estimations of requirements for recoverable spares were consistent with base stockage policies.

The early body of work at RAND exploring the demand modeling problem culminated in the work of Feeney and Sherbrooke. They dealt effectively with virtually every demand modeling problem raised by those whose work preceded theirs. Moreover, their interests extended beyond the demand modeling problem; they also made important contributions to supply stockage policy, at least for peacetime operating stocks. Feeney left RAND in 1966, but Sherbrooke continued to pursue important issues in stockage policy. In a 1966 RM [43], he discussed compound Poisson processes and their application to the demand modeling problem in anticipation of the need for a more tractable form of probability distribution for use in a multi-echelon inventory model called METRIC (Multi-Echelon Technique for Recoverable Item Control), which would have an important effect on inventory management policies and practices throughout the world, especially in the large inventory systems of military organizations.

¹A *compound Poisson process* is a stochastic process in which the number of arrivals that occur in a specified time period is described by the Poisson distribution, but with each arrival more than one event may occur. The number of events that occur with each arrival is described by a compounding distribution. In the present context, an arrival would be represented by the receipt of a requisition for spare parts; the quantity of parts requested on each requisition would be the compounding random variable. Thus the distribution of the number of parts requested in the time period would have a higher variance than the simple Poisson arrival process.

METRIC [44] extended the earlier work by Feeney and Sherbrooke to the multi-echelon base-depot inventory system. It assumed no lateral resupply and did not provide for condemnations. Its key to solution of the multi-echelon system was the application of Little's theorem [45] to the problem of estimating depot delay time, i.e., the time from receipt of a requisition by the depot to the shipment of the item, which of course is a function of the depot stock level. **METRIC** was described in laymen's terms in a subsequent publication [46]. The views of Feeney and Sherbrooke also shaped the aggregate stockage policy for EOQ items at base level suggested by Lu and Brooks [47] and published in June 1968. The publication of the **METRIC** development was also followed in June 1969 by an RM by Fishman, who suggested alternative methods of forecasting requirements for depot maintenance [48], in which, incidentally, he suggested the use of exponential smoothing in forecasting NRTS actions and clearly demonstrated its superiority to an eight-quarter moving average.

METRIC was eventually implemented by AFLC as the variable safety level (VSL) algorithm in 1975. It survived as the principal ingredient of safety stock computations until very recently when it was replaced by the Aircraft Availability Model (AAM), a **METRIC**-based computation that took explicit account of weapon system complexity in terms of the number of LRUs whose availability affected the availability of aircraft. Unfortunately, neither of these implementations took advantage of much of the knowledge that emerged from RAND's work in demand modeling. The mean demand rate was estimated with an eight-quarter moving average of worldwide demand counts. More will be said about the logical flaws of these procedures below. In short, **METRIC** found its way into the Air Force requirements system, but improved techniques for forecasting demand did not. Despite the problems associated with **METRIC**'s implementation, it was the seminal work in multi-echelon inventory theory, and it has had a profound influence on the course of inventory theoretic developments ever since.

THE LATER YEARS: 1969-1990

The work of Feeney and Sherbrooke was sufficiently satisfying, given RAND's views and interests at the time, that the topic of demand modeling was not reopened for more than a decade. The occasion for the renewed interest in demand modeling in the early 1980s was a shift in focus from the modeling of peacetime demands for aircraft spare parts to logistics support in wartime.

Among the issues that quickly emerged in the decade of the 1980s was that of the perturbation in demand processes induced by the transition from peacetime to wartime. Demands for resources were assumed to be directly proportional to flying hours (as they still are by the current system), and flying hours were anticipated to increase dramatically in wartime, especially in the early days of a war. This topic of nonstationarity in demand received considerable attention from RAND researchers, and led to several significant developments in capability assessment methodology and some advances in demand modeling. In the first formal RAND publication dealing with nonstationarity in demand in 1980, Muckstadt [49] developed an approximation to the probability distribution of the number of items in re-supply based on the assumption that the demand process was described by a nonstationary Poisson distribution and compared it to an exact derivation. The approximation was shown to be excellent. He also demonstrated that, in the face of nonstationarity in demand, it is important to model the nonstationarity explicitly, since assumptions of stationarity can lead to serious errors in the allocation of assets. He also discussed how the development could be applied to the initial provisioning problem.

Nonstationarity in demand was an important feature incorporated into a series of capability assessment models generically referred to as Dyna-METRIC. In its original development by Hillestad [50], Dyna-METRIC was a noteworthy departure from the traditional steady-state models of the time. Although the initial focus of RAND's research in wartime logistics support in the 1980s was on the transition from peacetime to wartime, the important advantage provided by models that explicitly accounted for nonstationarity was quickly put to good use in logistics policy analysis concerned with the more general problem of combat logistic support. A later version of Dyna-METRIC [51] was adopted by AFLC as its standard capability assessment model and incorporated in its Weapon System Management Information System (WSMIS). Pyles [52] suggested approaches to its use in practical logistics management applications.

These early versions of Dyna-METRIC were analytic models that were relatively efficient in running time but lacked some of the richness needed for evolving logistics policy studies at RAND. A later version called Dyna-METRIC Version 5 [53] was a hybrid analytic-simulation model in which constrained repair capacity and management adaptations such as cannibalization, lateral repair, and lateral supply could be explicitly represented. Unfortunately, Version 5 was not able to deal with the indenture relationships among LRUs and SRUs. A new version, Version 6, currently being used, corrects

that deficiency. All of these developments in capability assessment methodology were undertaken to pursue policy issues in logistics research at RAND. Version 4 was sufficiently tractable and useful that it found its way into AFLC's standard systems. Later versions may also be implemented in the future.

Some explorations of the mathematics associated with nonstationary demand processes were done more or less in parallel with the development of these several versions of Dyna-METRIC. Crawford [54] and Carrillo [55] published results that extended Palm's theorem [36] to nonstationary Poisson processes and to nonstationary compound Poisson processes, respectively. Crawford asserted that his result extended to the compound Poisson case but did not treat this more general problem explicitly. In an explanation for his approach, he made an observation that is especially pertinent in the context of the current work:

The practice of assuming an arrival process is compound Poisson when the data exhibit a variance-to-mean ratio greater than 1, as is often advocated in the literature, has little legitimate justification. If the variance-to-mean ratio is *significantly* larger than 1 (whatever that means), the arrival process is more often nonhomogeneous than [it] is compound Poisson. . . . In inventory applications, treating the arrival process as a compound Poisson when it is in fact nonhomogeneous Poisson can be shown to have a significant effect on the optimal stocking plan. . . .²

Crawford attributes the latter observation to J. Y. Lu, a former RAND researcher. It was never published. As we will show in discussion below, compound Poisson distributions are applicable to only a small proportion of the items in the recoverable inventory management system (the vast majority of requisitions being for quantities of one each), whereas a form of nonhomogeneous model seems almost universally applicable and, indeed, provides significant improvements in demand forecasting accuracy.

Crawford [2] provided a comprehensive description of the variability in demand for large sets of recoverable aircraft spare parts and pointed out the implications of that variability for readiness and sustainability and for spares requirements estimations and capability assessment modeling. He also showed that the numbers of items in resupply, especially in the depot repair pipeline, were not only highly

²A *nonstationary* stochastic process is simply one in which the arrival intensity varies over time. Throughout this report, we use the terms *nonhomogeneous* and *nonstationary* interchangeably, although in general usage nonstationary processes are a subset of nonhomogeneous processes.

variable but on average were higher than assumed in requirements estimations and capability assessment modeling. He also correctly pointed out that it is really the contents of the resupply pipeline and its variability, rather than variability in demand, to which system performance is most vulnerable.

This research by Crawford [2] and the research he had done immediately before it reopened many of the fundamental questions initially raised in the early years by Brown and others. Crawford pointed out, for example, that the demand models incorporated in current logistics management decision processes, most notably spares requirements estimation and capability assessment modeling, are inconsistent with observed data; moreover, the inconsistencies suggest that we really don't understand the demand process as well as we seem to imply in our characterizations of the nature of the process, its variability, its stationarity, or its relationship to activity level (e.g., flying hours). It is important to note for our later discussions that Crawford's work in this case examined worldwide aggregations of demands partitioned quarterly. This fact shaped his results in important ways, as we shall see below.

Crawford's work, which was directed toward better understanding of the behavior of demand processes for aircraft spare parts in peacetime, coupled with RAND's interest in logistic support in wartime, provided much of the motivation for an important body of work done at RAND in the 1980s, most of it in a major undertaking of Project AIR FORCE, which became popularly known as the Uncertainty Project.³ The very title of this work reflected the thrust of Crawford's analyses and the thinking of others at RAND that the variability in peacetime demands for aircraft spare parts would be compounded in wartime by system disruptions, resource losses, and the inevitable surprises of the combat scenario. In the course of the Uncertainty Project, a set of management initiatives emerged that was intended to mitigate the effects of uncertainties in resource demands. Those initiatives were given the name CLOUT (Coupling Logistics to Operations to meet Uncertainties and the Threat). The CLOUT initiatives are described by Cohen, Abell, and Lippiatt [20]. The following discussion draws heavily from the Summary of their report.

³The Uncertainty Project was part of RAND's Resource Management Program. It was formally entitled "Enhancing the Integration and Responsiveness of the Logistics Support System to Meet Peacetime and Wartime Uncertainties." It was sponsored by Headquarters, USAF/LEX, and Headquarters, AFLC/XP.

The CLOUT initiatives are examples from a taxonomy of more generic strategies for coping with uncertainty described by Hodges [56, pp. 20-24]. The uncertainty is of essentially two kinds: (a) *statistical* uncertainty, defined as variability observed in repeatable phenomena, and (b) *state-of-the-world* uncertainty, defined as uncertainty that arises in phenomena that are not repeatable, not observed or observable, or both [56, pp. 8-14]. Planning for wartime is fraught with uncertainty of the latter kind; in fact, state-of-the-world uncertainty can fairly be said to dominate the wartime scenario as well as military planning for it.

Essentially, the CLOUT initiatives generally place less reliance on a richness of spares and take greater advantage of more flexible resources such as maintenance and distribution. That strategy derives logically from the difficulty and cost of a "buyout" strategy that would attempt to provide ample quantities of spares, for example, to allow the system to cope with the levels of uncertainty in demand that it might face, especially in wartime. At the theater level, there are significant payoffs to be gained from alternative operating policies for theater distribution systems that take fuller advantage of responsive lateral resupply and lateral repair options. They also suggest that closer coupling of the depot repair system to the combat forces has significant payoff in aircraft availability. The thrust of the thinking underlying CLOUT is to rely less on an ampleness of goods and more on management adaptations. (Hodges describes these as examples of passive and active strategies, respectively.) That thinking has important implications for system design as well as management. Many characteristics of the current system need to be changed to achieve the kind of relevant, timely, and robust performance needed to cope with unanticipated, urgent demands for resources.

The magnitude and pervasiveness of wartime and peacetime uncertainties also suggest the need in both capability assessment models and spares requirements estimation to take explicit account of realistic levels of peacetime and wartime uncertainty and the ability of the support system to cope with them through the kinds of adaptive behavior exemplified by the CLOUT initiatives. In the face of such uncertainties, the robustness of system performance in a variety of scenarios is at least as important as the level of system performance in a single, specified planning scenario.

The implications of this approach for logistics management, logistics policy analysis, and the design of logistics management systems are fundamentally important. In problems involving state-of-the-world uncertainty, analysts have traditionally retreated to analytic methods

intended to deal with statistical uncertainty, methods with which they are familiar and comfortable. Thinking clearly about the implications of state-of-the-world uncertainty for particular applications is both difficult and unfamiliar for many persons involved in policy analysis. Perhaps the most important central message of this work is the need to take explicit account of uncertainty, particularly state-of-the-world uncertainty, in formulating policies and designing systems, and to take explicit steps to ensure that the performance of those policies and systems is robust in the face of those uncertainties.

The CLOUT initiatives have been incorporated into the Air Force's new Logistics Concept of Operations. They clearly have important implications for spares and repair requirements estimation. Those implications are the principal interest of the larger project of which this demand forecasting and modeling work is one part. Our review of past RAND research in demand forecasting has presumably made it clear that whatever improvements might be possible in demand forecasting should be fully exploited, even in the face of extensive employment of management adaptations in logistics operations. Abell et al. [57] describe, for example, a depot-level repair prioritization and asset allocation mechanism which is intended to be much more adaptive in the face of uncertain demands than are the current depot repair management and stock control and distribution systems. As adaptive as it is, it still depends to some extent on demand forecasts. The need for improvement in our ability to forecast was recently made emphatically clear by Lippiatt [1], who observed large errors in forecasting NRTS actions using the traditional methods of AFLC's standard system. He observed that the further into the future the forecast was made, the greater the error, perhaps an unsurprising result in retrospect, but the error was consistently positively biased, a phenomenon that may have been associated with the time period examined.

In spares procurement, too, because of the long lead times frequently encountered, accurate demand forecasts play an important role in shaping the mix of spares procurement actions to achieve an effective stockage posture. Thus our exploration of demand forecasting and modeling is not only an important part of our requirements estimation research; it is also important to a variety of logistics management decisionmaking. We believe it can improve such decisionmaking, especially over the longer planning horizons typically associated with recoverable spares procurements, and enhance the Air Force's ability to achieve greater effectiveness from its available resources. We now turn our attention to the Air Force's approach to demand modeling incorporated in the current system for estimating spares and repair

requirements before describing in following sections more effective alternatives to that approach. A more complete description of the current recoverable spares requirements system can be found in Ref. 58.

DEMAND MODELING IN THE CURRENT SYSTEM

A traditional approach is taken to modeling demand by AFMC's current spares and repair requirements estimation system. The worldwide mean demand rate for each master stock number is estimated in a straightforward way, item by item, using an eight-quarter moving average.⁴ Because of data-collection, transmission, and processing times, the demand counts available in the central system are at least one quarter behind. Thus, if one were at a point in time in the middle of quarter n , one would lack visibility of the demands to date in quarter n and the demands that occurred in quarter $n - 1$. The demands in quarter $n - 1$ typically become available in the central system late in quarter n .⁵

The eight-quarter moving average gives as much weight or importance to demands eight quarters in the past as to demands only one quarter in the past. The moving average does have the advantage, however, of smoothing the demands across quarters so that the system's estimate of demand rate is not so volatile as it would be using less data. On the other hand, it is not as responsive to changes in demand rates as some alternative approaches. The current system assumes that demand rates per flying hour are constant; i.e., demands are directly proportional to flying hours. Thus if flying hours are projected to increase by x percent, expected demands will also increase by x percent. One thrust of this report is to present alternatives to these demand estimating techniques that substantially reduce expected forecasting error.

The requirements for safety stock are computed by the current system using an estimate of the probability distribution of the numbers of items of each master stock number family in resupply based on steady-state assumptions, i.e., stationary demand rates, known activity levels, constant pipeline times, and constant NRTS rates. The negative binomial probability distribution is used to describe this

⁴The eight-quarter moving average has the form $\sum D_i / \sum H_i$, where the (D_i) are the demands observed over eight quarters, the (H_i) are the item flying hours derived from all of the applications of the item, and the summations are taken over the eight quarters of past experience.

⁵Although the D035C system has very current demand data available, the D041 system, the system that supports requirements estimations, does not.

random variable. The VTMR of the distribution is based on empirical observations by Stevens and Hill [59] suggesting that the observed VTMR of demands for recoverable spare parts was an increasing function of the mean demand rate, but increasing at a decreasing rate. The observed VTMR is erroneously treated as though it were the true VTMR. It has the form of the unbiased estimator of the variance divided by the observed mean demand rate. Stevens and Hill chose a power function on empirical grounds to describe the relationship of the form $VTMR = ax^b$, where x represented the mean demand rate. At the time of this writing, the variable x represents the expected number of items of a given type in resupply, i.e., the item pipeline, and the values of a and b are 1.132477 and 0.3407513, respectively. If the resulting VTMR is less than 1.01, it is set equal to 1.01; if it is greater than 5.0, it is set equal to 5.0.

A priori, there is no legitimate basis for the assumption that the VTMR should vary in any systematic way with mean demand. Yet, the data, as aggregated by Stevens and Hill, do suggest such a relationship.

It is not clear that the power function used in the current system is appropriate. The actual fit between calculated item pipelines and observed values of the VTMRs assigned to the probability distributions of those pipelines at the individual item level is remarkably poor. Figures 2.1 and 2.2 illustrate this contention. The observations simply do not fit the data, probably because of nonstationarity in the underlying demand process that, as we will show below, can induce a profound overestimation of the VTMR. The use of the power function is not on solid ground given what we have learned about the numerical values of the VTMR estimator in the presence of nonstationarity in demand; nevertheless, it is difficult to find a better alternative approach, although the numerical parameters of the function can be improved by accounting explicitly for forecasting uncertainty, which the current method ignores.

Our comments about the current system's approach should, perhaps, be more forgiving because the problem of sensibly estimating the variance-to-mean ratio of the distribution of the number of items of each type in resupply is very difficult. Sherbrooke [44, p. 32] and, later, Hodges [56, p. 32] suggested that numerically estimated values of the VTMR could be computed using the method of maximum likelihood; however, this approach is traditionally applied on an item-by-item basis. The numerical values that result from such an approach are highly volatile and would probably result in highly volatile estimates of the total number of spares of each type the requirements

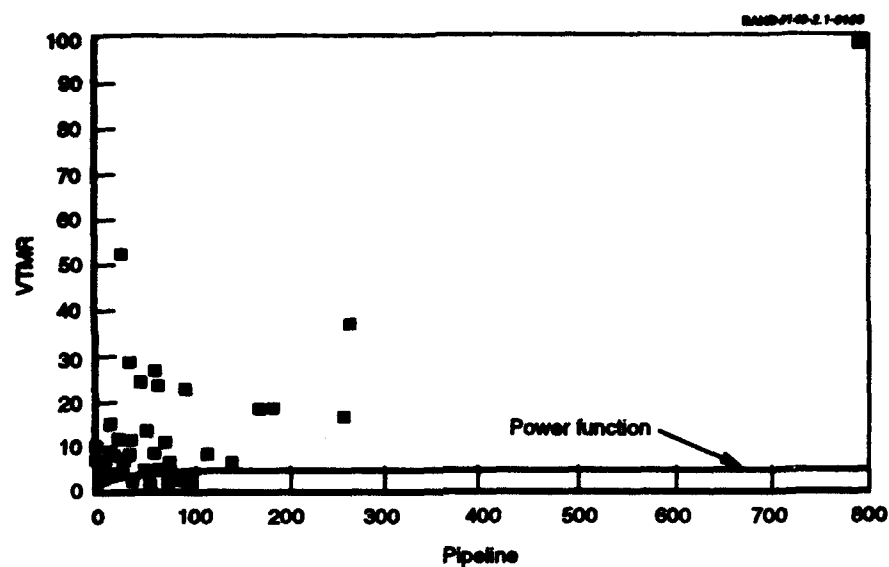


Figure 2.1—Observed VTMR as a Function of Expected Number in Resupply

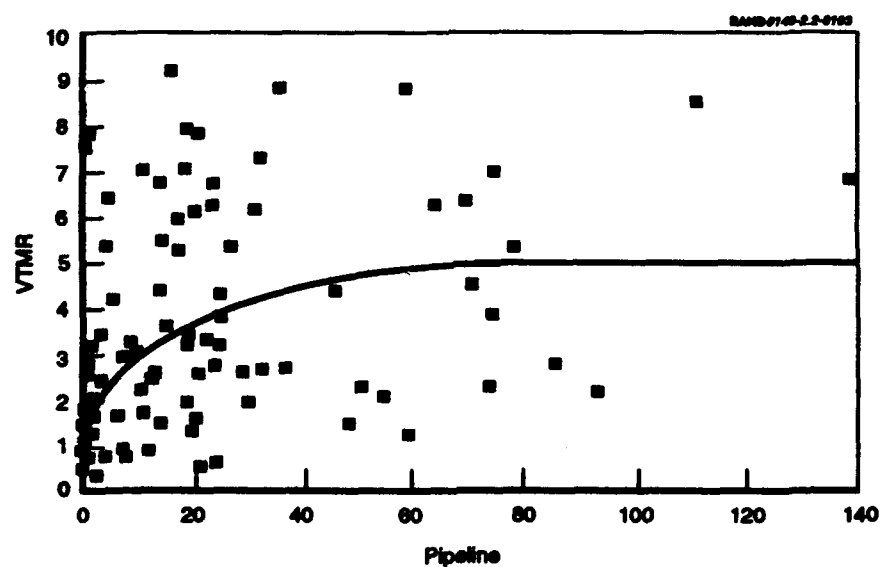


Figure 2.2—Enlargement of Lower Left Portion of Figure 2.1

system estimates should be in the inventory system. Such volatility is one of the principal causes of items being in long supply. Thus an approach to estimating the variance of the probability distribution of the number of items in resupply that is based in some way on aggregations of items or that employs some kind of smoothing, stabilizing, or filtering technique seems to be indicated. Alternatives to the current system's method of establishing the variance were evaluated in this research, and we recommend one that represents a modest improvement to the current approach, but one that is on much more solid logical ground. The discussions in Section 6 shed additional light on the problem of variance estimation.

Although the principal thrust of our research in the estimation of requirements for aircraft recoverable spares and depot repair is to bring explicit recognition of flexibility and responsiveness to the spares and repair requirements estimation process, in this report we specifically address issues of demand modeling and forecasting in the hope of enhancing the cost-effectiveness of the estimation process.

3. SOME EXPLORATIONS OF BASE-LEVEL DEMAND DATA

Prompted by our increasing interest in the implications of uncertainty for effective combat logistics support, and reinforced by the findings of Crawford [2] that line items with high demand variability tend to be the troublesome ones in terms of inventory system performance, we reopened some of the demand modeling issues originally raised by the examination of base-level demand data by earlier RAND researchers. The data explorations described in this section led to two principal findings, each of which is important to the Air Force's ability to provision itself cost-effectively with aircraft recoverable spare parts: (a) There is strong evidence that parts demand processes are, in general, nonstationary, and (b) certain characteristics of specific line items which are, in many cases, known in advance tend to be associated with high levels of demand variability. The implications of the first of these observations are important for modeling the process correctly, thus improving our ability to forecast demands and to understand the demand process somewhat better than we now do. The second observation is important to our ability to manage logistics support better, to know when exceptional management action is needed, and to improve our formulation and implementation of sensible policies for spares investments and inventory system management.

The Air Force's Logistics Management Center (LMC) at Gunter AFB, Alabama, kindly provided RAND with data from several bases from which we were able to infer actual demands on base supply by base maintenance activities. One of the bases from which data were provided was Bitburg Air Base, FRG. The Air Force has, for some time, maintained a wing of F-15C/D aircraft at Bitburg, and we acquired 48 weeks of transaction data from Bitburg from which we extracted transactions, then demands, for F-15 recoverable LRUs and SRUs. The observations that follow were made from those data.

ITEM CHARACTERISTICS THAT TEND TO PREDICT HIGH VTMRs

In the Bitburg analysis, we were able to observe monthly demand rates per flying hour as well as demands independent of flying hours. The flying hour production during the 48-week period to which the data applied was sufficiently regular that it did not make a substantial difference to these findings; however, the data described only

monthly flying hour activity, not daily flying hours. In comparing the distribution of VTMRs among line items using monthly demands and monthly flying hours, it was clear that the use of flying hours was not especially important because they did not vary much. Since much of what we learned derived from our ability to observe demands on every line item every day, we chose to ignore demand rates per flying hour in favor of analyzing simple counts of demands. It must suffice here to say that we are confident that our observations would be as valid had we accounted for flying hours at every step of the analysis as they are without having done so.

Items with large QPAs (quantities per application) or action quantities greater than 1 tended to have high VTMRs. These two subsets of line items had a substantial intersection. In the case of a demand transaction, the action quantity is simply the number of units of issue (e.g., each, dozen, pair, box) reported in the transaction.

In the discussion that follows, we divide the demand data into 24 two-week intervals and, for line item j , compute the observed mean demand rate, m_j , as the total number of demands in the 24 two-week periods for line item j divided by 24. The VTMR estimator, r_j , is simply the unbiased estimator of the variance divided by m_j .¹ There were 973 items in the sample, all F-15 recoverable LRUs and SRUs. Demands that were satisfied by intermediate-level repair, that is, by maintenance itself, were excluded from this sample. The remaining demands are those satisfied by the inventory system, not by repair. The distribution of $\{r_j\}$ reflected 75 percent above 0.96, 50 percent above 1.03, 25 percent above 1.61, and 5.2 percent above 3.0. The distribution had a very long tail with some extraordinarily high VTMRs, the highest being a startling 56.16. The unweighted mean of the VTMRs was 1.79 with a remarkably high standard deviation of 3.52. Eliminating all line items with action quantities greater than 1 or QPAs greater than 4 reduced the sample size from 973 to 918 and dramatically altered the distribution of $\{r_j\}$. Its unweighted mean was reduced from 1.79 to 1.34 and its standard deviation from 3.52 to 1.38. The number of line items with VTMRs greater than 3.0 decreased from 51 to 16, and the number of line items with VTMRs greater than 5.0 decreased from 30 to 5. Thus, of the 55 line items eliminated because of large QPAs or action quantities greater than 1, 35 had VTMRs greater than 3.0 and 25 had VTMRs greater than 5.0. In percentage terms, 5.2 percent of the original 973 line items had VTMRs greater than 3.0, and 3.1 percent had VTMRs greater than

¹The unbiased estimator of the variance is given by $[n/(n-1)] [\sum x_i^2/n - (\sum x_i/n)^2]$, where x_i is the number of demands in the i th two-week period and $n = 24$.

5.0. These percentages decreased to 1.7 and 0.5, respectively, of the 918 line items remaining after removing those with action quantities greater than 1 or QPAs greater than 4.

Several observations emerge from these results. Most important, perhaps, is the fact that the distribution of observed VTMRs (we are talking about VTMR estimates of the form previously described as r_j throughout this discussion) at base level based on a two-week partitioning of the demand data is dramatically different from that at systemwide level as reflected in the spares requirements database (eight quarterly observations of worldwide demands). As we will discuss below, our partitioning of the data into two-week intervals in contrast to the quarterly intervals used in the spares requirements system accounts for some of this difference, but not all of it. The difference is probably amplified by the fact that the Bitburg data span 48 weeks whereas the systemwide VTMRs are computed using eight quarters of data. The implications of this observation for demand modeling are important: (a) One should not observe numerical values of systemwide VTMRs and apply them directly to base-level demand processes as the current spares requirements system does, (b) one cannot infer much about the volatility, uncertainty, or unpredictability of base-level demands from systemwide data aggregated quarterly, especially if the partitioning is different, and (c) there are serious measurement problems associated with the VTMR estimator.

A second observation is that action quantity (in this case the number of parts requested in a requisition) acts as a VTMR multiplier. If a particular stochastic process is simple Poisson, then its VTMR is 1. If, with each requisition arrival, n demands occur, then the VTMR of the demand process is n , but the process is no longer simple Poisson; it is compound Poisson (in this case with a degenerate compounding distribution). Thus, comparing its VTMR to 1 may tend to lead one to conclude that the process is somehow out of control, wild, or unpredictable, based on the value of n , when in fact it is exactly as predictable as a simple Poisson process. Thus, the requisitioning of more than one unit at a time of some line items causes the VTMRs of their demands to be greater than the VTMRs of their requisition arrivals. It is important to note that in the Bitburg data only 3.7 percent of the original 973 line items (36 line items) reflected any action quantities greater than 1; thus Crawford's and Lu's observations about the unsuitability of compound Poisson probability distributions to describe parts demand processes are strongly reinforced by the Bitburg data. Recall our earlier quote from Crawford [54, p. 32]:

The practice of assuming an arrival process is compound Poisson when the data exhibit a variance-to-mean ratio greater than 1, as is often advocated in the literature, has little legitimate justification. If the variance-to-mean ratio is *significantly* larger than 1 (whatever that means), the arrival process is more often nonhomogeneous than [it] is compound Poisson. . . . In inventory applications, treating the arrival process as a compound Poisson when it is in fact nonhomogeneous Poisson can be shown to have a significant effect on the optimal stocking plan. . . .

Since our analysis of demand data from Bitburg was confined to demands for recoverable items, it is not surprising that so few line items experienced action quantities greater than 1; the Air Force's reorder policy for recoverable items implemented in the standard base supply system is $(s, s - 1)$; i.e., when the number of assets on hand plus due-in minus due-out falls below the stock level, s , an order is placed immediately (usually daily, as a practical matter) for replenishment. This is commonly described as a continuous review reorder policy with an order quantity of 1. For items with action quantities greater than 1, the compound Poisson model may be appropriate to accommodate the higher variance that tends to be associated with these items. A large QPA does not necessarily imply a compound demand distribution. All we know from the Bitburg data is that large QPAs tend to be associated with high demand variability. As we pointed out above, the set of items with action quantities greater than 1 and the set with large QPAs overlap significantly. Of the 36 line items reflecting action quantities greater than 1 in the Bitburg data, only six, i.e., 16.7 percent, were known to have a QPA of 1.² These observations about action quantities simply suggest that large VTMRs do not necessarily imply unpredictability.

After eliminating the items with action quantities greater than 1 and QPAs greater than 4 from the distribution of observed VTMRs, $\{r_j\}$, there were still five of the remaining 918 line items with VTMRs greater than 5.0, the upper bound established for VTMRs used in the spares requirements estimation process to describe the variability of the number of assets of a given line item in resupply. Their VTMRs were 5.5, 6.5, 7.2, 18.7, and 34.6. Their QPAs were 1, 3, 3, 4, and 1, respectively. The first three were SRUs, the latter two LRUs.

²The QPA (quantity per application) of an item is the quantity of the item installed in the next higher assembly. We hasten to add for the sake of completeness that the QPAs of two of the 36 items were shown in the D041 application file as 0. One of them was an engine vane, very likely an item with a large QPA; the other was an expensive switch in the 5841 property class whose true QPA is still unknown.

The highest VTMR was associated with the nose landing gear strut, a \$26,500 component which, along with one of its principal SRUs, the \$15,600 nose landing gear piston, exhibited rather remarkable demand behavior. The demands for the strut and piston observed in each of the 24 two-week periods were:

Strut	1	2	2	2	0	4	2	0	1	0	0	1	1	0	1	0	0	1	2	1	51	0	0	1
Piston	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	55	0	0	63

It is interesting to note that the demands for the large quantities of both items, 51, 55, and 63, were all backordered with no corresponding MICAP event, although over the entire period, there were 10 MICAPs reported on the strut but none on the piston. Our assumption about this wildness is that it is the result of some kind of policy intervention, perhaps a time change replacement, perhaps an intervention of the IM or SPM, compliance with a technical directive, or a safety-of-flight consideration.³

It is not surprising that three of these line items with high VTMRs (6.5, 7.2, and 18.7) have QPAs greater than 1 (3, 3, and 4). The item with the VTMR of 18.7, an \$1,131 SRU in the 1440 federal stock class, experienced 186 demands during the 48 weeks, not one of which was satisfied through an issue by base supply. Of the 186 demands, 86 were satisfied by WRSK withdrawals and the remaining 100 were backordered. No MICAP was ever reported as a result of shortages of the item; it had no apparent impact on mission capability except that it presumably depleted the WRSK.

The remaining line item hardly deserves attention. It experienced only nine demands in the 24 two-week periods. It has such a high observed VTMR because seven of the demands occurred in one period. All were satisfied by off-the-shelf issues.

THE VARIANCE OF THE ESTIMATOR, r_j

One question that is prompted by an investigation like this is the role of chance in producing particular values of observed VTMRs given some numerical value of the VTMR underlying a stochastic process. This question is especially relevant to the present discussion simply because the variance of the VTMR estimator, r_j , is quite large. A VTMR of, say, 1.5, coupled with a mean demand rate of, say, 0.5 de-

³If this was the case, the peak in demands would have been a one-time occurrence and should not have been reflected in estimation of the item's future demand rate.

mands per period over eight periods, is likely to yield an observed VTMR greater than 3.0 about 3 percent of the time.

It may provide some intuition about these processes to recall that, in the Bitburg data, about 5.2 percent of the 973 original observed VTMRs were greater than 3.0. After removing the 55 line items with action quantities greater than 1 or QPAs greater than 4, only about 1.7 percent of the 918 remaining observed VTMRs were greater than 3.0. For these 918 items, the distribution of observed VTMRs is perfectly consistent with an assumption that all the parts have an underlying VTMR of 1.5 or less.

It is clear from our observations of large numbers of random realizations of these demand processes as well as the theoretical work by Hodges [60] that the error distribution of the VTMR estimator itself bears much of the responsibility for the distributions of observed VTMRs we see in parts demand processes. The estimator is also very sensitive to nonstationarity which, as we discuss in Section 6, also contributes significantly to shaping the observed VTMRs.

NONSTATIONARITY IN BASE-LEVEL DEMAND PROCESSES

In our analysis of the Bitburg data, we tried partitioning the 48 weeks of data in several ways: 24 two-week periods, 12 four-week, 8 six-week, 6 eight-week, and 4 12-week periods. We found a remarkably systematic relationship between the observed VTMRs and the coarseness of the partitioning. Tables 3.1 and 3.2 show the results for all parts and for the subset of 918 with QPAs less than 5 and action quantities of 1.

If a process is stationary, the expectation of the observed VTMR is independent of the partitioning, although its variance increases with the coarseness of the partitioning. In the stationary case, the VTMRs in Tables 3.1 and 3.2 would be approximately equal within each table, independent of the partitioning. In the actual case being examined here, where we pool the results of alternative partitionings of over 900 random realizations of demand processes and observe such systematic variation, it seems very likely that nonstationarity is involved, owing simply to the lack of any other explanation. While this systematic variation does not prove nonstationarity, it strongly suggests it.⁴

⁴Excluding items with action quantities greater than 1 or QPAs greater than 4 lowers the values shown in Table 3.1, but the pattern remains largely the same as shown in Table 3.2.

Table 3.1
Effects of Partitioning on Observed VTMRs
(all 973 parts)

No. of Periods	Weeks in Period	Avg. Observed VTMR	95 Percent Confidence Interval
24	2	1.79	(1.68, 1.90)
12	4	1.91	(1.68, 2.16)
8	6	1.92	(1.68, 2.16)
6	8	2.11	(1.76, 2.46)
4	12	2.20	(1.85, 2.55)

Table 3.2
Effects of Partitioning on Observed VTMRs
(918 parts with QPAs less than 5 and action
quantities of 1)

No. of Periods	Weeks in Period	Avg. Observed VTMR	95 Percent Confidence Interval
24	2	1.35	(1.26, 1.44)
12	4	1.41	(1.32, 1.50)
8	6	1.45	(1.35, 1.55)
6	8	1.49	(1.39, 1.59)
4	12	1.60	(1.47, 1.73)

The demand modeling methods that we will discuss in the sections that follow derive much of their power from the fact that they explicitly model the nonstationarity in the demand process and separate its effects from those of the inherent variability of the process around its nonstationary location parameter. Before introducing these alternative methods, let us first reflect briefly on the implications of these discussions.

IMPLICATIONS OF THE BASE-LEVEL ANALYSES

One may wonder whether we can prudently infer as much as we have inferred from data collected at a single base on a single weapon system. Such a concern is well founded and suggests the need for extensions to these analyses.

Our principal observations inferred from the Bitburg data are:

- The observed VTMRs associated with parts demands at base level are very different from those of quarterly demands aggregated

worldwide because demands per flying hour tend to be positively correlated across bases.

- Large QPAs and order quantities greater than 1 contribute significantly to high observed VTMRs.
- The variance of the VTMR estimator explains to a substantial extent the occurrence of large values of observed VTMRs.
- Parts demand processes appear, in general, to be nonhomogeneous rather than compound processes.

It is important that we distinguish our inferences about demand uncertainty on the basis of worldwide aggregations of quarterly demands, for example, from those that we make from observing base-level demand processes. We caution the reader that our comparison of the distributions of VTMRs at the base and system levels here is flawed by the fact that the base-level data span only 48 weeks, whereas the worldwide data in the spares requirements database span two years; the periods of aggregation are different as well, as we have pointed out repeatedly. The levels of variability implicit in distributions of observed VTMRs are quite different for the two echelons. It is also important to note that the item characteristics (QPA and order quantity) that help explain high observed VTMRs at base level may also be helpful in explaining high VTMRs systemwide. Quarterly partitioning of demands aggregated worldwide can produce very different distributions of observed VTMRs from those observed at base level because bases are not independent, their covariance terms typically being positive. The demand processes of many parts are nonstationary, so if the worldwide data are partitioned quarterly and the base-level data are partitioned more finely, the worldwide data will exhibit higher VTMRs.

We make the observation that these demand processes tend, in general, to be nonhomogeneous rather than compound processes because of the sensitivity of the distribution of observed VTMRs to the coarseness of partitioning of the demand stream, and because of the small number of requisitions with order quantities greater than 1. This observation itself is important to our efforts to develop more rational models of the demand process that will yield improved demand forecasts and more effective investments in the face of resource constraints. In the several sections that follow, the effectiveness of taking explicit account of nonstationarity is clearly demonstrated.

WARTIME UNCERTAINTIES

It is important to emphasize again that the observations discussed here apply to peacetime demand data, not necessarily to demand processes in wartime, not because the underlying failure process changes, necessarily, but because the mix of demands the system sees from units engaged in combat is likely to look quite different from peacetime demands for many reasons. Our uncertainty about peacetime demands is likely to be compounded in wartime in ways that we cannot readily foresee. Many events that are essentially unknowable in advance are more likely to occur in wartime and generate demands to which the logistics system must be able to respond quickly.

The notion of state-of-the-world uncertainty and its distinction from statistical uncertainty is the kernel of the issue here. Most of the discussion in this report pertains to uncertainty of the statistical variety. While admittedly important to effective and efficient peacetime support, we do not fully understand the implications of state-of-the-world uncertainty for wartime support. The problem of better understanding those implications for combat logistics support is undoubtedly important and deserves further research.

4. ALTERNATIVE APPROACHES TO DEMAND FORECASTING

As we saw in our review of past RAND research on the topic of modeling the demands for aircraft spare parts, the early researchers tended to think of these demand processes as stationary, i.e., as being random observations about an unknown constant. Recall that there were numerous references to the idea of "the true mean" in that early literature. The ideas we will introduce in the discussion that follows represent a different way of thinking about these stochastic processes, as time-varying or nonstationary. Nonstationarity (or nonhomogeneity) can produce large variances (therefore large VTMRs) even when the process at any given time or at any given base might be simple Poisson around a changing mean.

Any omission of an important explanatory variable in the demand modeling process results in variability in the mean being lumped with the variability about the mean. As a practical matter, it might be too costly to collect data related to every possible explanatory variable; thus one postulates a reasonable failure model and collects data related only to the variables in that model. Many potential explanatory variables are not explicitly considered. It is important to note, too, that data collection itself may contribute significantly to the variance we observe in the process through errors in collection, transmission, or processing. As we saw in Section 3, the design of the data-collection and reporting system may itself induce higher observed variability in demand owing to the constraint it places on our ability to partition the observations more finely. Thus the variability we observe in the demand process may have several components that are not explicitly identifiable. While we may not be able to isolate the contribution of every factor to the variance we observe in the demand process, there are gains to be made from separating the variance inherent in the demand process from that in its underlying location parameter. The precise functional form of the relationship between demands and flying hours as well as certain other explanatory variables is discussed in Appendix B.

The spare parts demand forecasting literature is replete with ideas for using other measures of activity to predict demands. Changes in the Air Force's data systems might also facilitate improved demand forecasting. In addition, many authors have looked at parts failures and found them to be readily explicable in retrospect (retrodiction being an easier task than prediction). The problem is that, even if we

understood the effects of many kinds of events on the demand process, such understanding isn't useful unless we can predict the events themselves. It is important to account explicitly for uncertainty as much as possible. By this we do not mean to collect every possible explanatory variable or put a stress measurement device on every part. We mean to incorporate a realistic assessment of our uncertainty. The alternative forecasting methods we discuss in this section facilitate explication of our uncertainty.

It may be helpful to try to characterize the nature and sources of uncertainty in the demand process. So far we have loosely lumped together nonstationarity, stochastic variation, and omitted explanatory variables. In the discussion that follows, we identify types of uncertainty that may affect the demand process: exchangeability uncertainty, model and structural uncertainty, data uncertainty, and coefficient and stochastic variance.¹

EXCHANGEABILITY UNCERTAINTY

The exchangeability judgment is a critical factor in the application of any forecasting method. To get some clear intuition about this judgment, we must mentally separate out some of the other sources of uncertainty. Consider a model that is perfect in the sense that we know all the coefficients and relationships exactly, without the uncertainty of statistical estimation. The key question is whether such a model would be a good model of the demand process in the future. If we believe the future Air Force will be much like today's Air Force with similar support structures, policies, and strategies, our judgment may be that our current model of the demand process may be exchangeable with a model of the demand process in the future. On the other hand, the Air Force might adopt a support structure with only two levels of maintenance or introduce technological change that could substantially affect the repairable generation process. Responses to exogenous shocks to the system such as might be felt in wartime could also induce changes in operating policies or procedures that could, in turn, affect the demand process. In such cases, our exchangeability uncertainty would be substantial and some explicit adjustment to our forecasting models might be needed.

Although reassuring ourselves about the exchangeability is intrinsically difficult, there are some things we can do to try to enhance exchangeability. When we build models, we try to include variables

¹Several of the ideas presented in this section are from our RAND colleagues David Draper and James S. Hodges.

that capture important differences between time periods. For example, we would want to include flying hours as an explanation of demand even if flying hours don't vary much in the short term because we may need to forecast demands in periods of increased activity. We do this because we feel more comfortable about the application of the model to different situations if the model includes variables that we think will explain the difference between the present and the future.

When exchangeability uncertainty is large, we tend to think of it as state-of-the-world uncertainty [56]. When it is small, we hope to accommodate it, at least crudely or partially, with the kinds of models we discuss in this section.

MODEL AND STRUCTURAL UNCERTAINTY

We will refer to the three kinds of model uncertainty as variable selection, variable specification, and structural uncertainty.

Variable selection uncertainty is the uncertainty introduced by our inability to know exactly and exhaustively which variables should be used to forecast the demands for a particular part. This would include the uncertainty from our inability to use certain desirable variables because of lack of data on those variables.

Variable specification uncertainty is our uncertainty about exactly how each variable should be incorporated or specified in the model. For example, should flying hours be transformed into quadratic or logarithmic form or incorporated linearly?

Structural uncertainty is the uncertainty in how various relationships interact to produce the observed outcomes. For example, does the NRTS rate depend on the demand rate? Do the bases NRTS more parts during periods of stress? Rule-versus-model uncertainty, a special case of structural uncertainty, addresses whether certain rules or interventions take place that are not captured by the model. For example, if the Air Force changed its transportation policies in a way that dramatically increased item pipelines, they would differ markedly from what our models suggest.

DATA UNCERTAINTY

Data uncertainty is the uncertainty that results from errors in recording, transmitting, or processing data. It is the uncertainty that results from lost magnetic tapes and corrupted data. It is the consequence of reporting omissions that may well be the result of necessary circumventions of standard procedure. As is the case with the omis-

sion of important explanatory variables, this is a difficult kind of variability to incorporate only if it changes from period to period. For example, a systematic under-reporting of flying hours by 10 percent does not present much of a problem for demand forecasting. Presumably the system will just end up with a demand rate that is higher than the "true" demand rate. This will result in more or less the appropriate decisions. The problem is when the data errors vary from period to period, inducing additional, inexplicable variability.

PARAMETER OR COEFFICIENT VARIANCE

Another source of uncertainty is coefficient variance. This is the uncertainty that results from estimating coefficients in a model statistically as opposed to knowing them exactly. For example, in the traditional regression model, $y = a + bx + \varepsilon$, coefficient variance would be the consequence of our not knowing the numerical value of b a priori. In many applications, this is the only source of uncertainty that is explicitly considered and, ironically, it is often the least important.

STOCHASTIC VARIANCE

The final source of uncertainty is the stochastic uncertainty resulting from a prediction. Even if the probability of a coin coming up heads is known to be exactly $1/2$, there is stochastic uncertainty about the outcome of any individual toss. In the regression modeling context, this corresponds to another draw from the error distribution (the epsilon in the expression $y = a + bx + \varepsilon$) for a future observation. The combination of parameter and stochastic variance is sometimes called predictive variance. The current system ignores parameter variance and all other sources of variability, incorporating only an estimate of stochastic variance as its representation of uncertainty.

ACCOUNTING FOR UNCERTAINTIES

The sum of all these sources of variability, except the traditional parameter and stochastic variance, makes up the nonstationarity that we see in parts demand processes. The bulk of it results from real causes that we do not understand a priori or do not measure. Although we cannot include all of these sources of uncertainty explicitly in our model of the demand process, we can attempt to quantify the magnitude of their variability in our models and decisions. As we will show, forecasting approaches that try to account for as many of these sources of uncertainty as practicable are more effective than the approach taken by the current system.

Our ability to account for these sources of variability depends upon the sum of their contributions to the variance of the process remaining roughly constant over time. This is an exchangeability judgment in the variance scale. It is possible, for example, that the mean of a process is drifting over time but the variance of that movement is constant. The models we consider in this section incorporate this logic, although it is by no means well founded. For example, there is every reason to believe that demand variability will increase radically during wartime. Our principal argument is simply that any attempt to incorporate nonstationarity is probably better than assuming that it does not exist; however, trying to incorporate nonstationarity in the variance adds a layer of complexity that is probably unwarranted.

An added bonus from incorporating nonstationarity explicitly in our models of the demand process is that we then have a handy knob to turn when we face future uncertainty that is larger than usual. For example, in the transition to wartime we could increase our estimate of the sum of these variance components to make the system more responsive to the first incoming wartime data. We may not be able to forecast the direction of change in mean demand rate in a particular set of circumstances, but we may be able to increase our estimate of the variance fairly easily to represent our increased uncertainty about the future. We should do everything of this kind that we can before we throw up our hands and claim that no forecasting of the future is possible.

THE KALMAN FILTER

We begin our discussion of alternative approaches to demand forecasting with the Kalman filter, a model that explicitly incorporates nonstationarity by treating the mean of the demand process as a random variable. The mean need not be a simple random variable with the Kalman filter; it might also be a somewhat more complicated model akin to a description of the mean as in a regression function. In this section we will explore the logic underlying the Kalman filter. Even the simplest Kalman filters have two important characteristics that make them very powerful in representing the particular and special features of parts demand processes: (a) They explicitly model the location parameter, or mean, of the process as a random variable, and (b) in exactly the manner suggested by Kenneth Arrow [28], they not only give greater weight to more recent observations than to older ones, they also assign greater uncertainty to forecasts in the distant future than to those in shorter planning horizons, an intuitively appealing notion.

Kalman filters are an especially rich class of models originally introduced in the electrical engineering literature by Kalman and Bucy [61]. Hence the name *filter* as opposed to the more colloquial *predictor*. They have since grown in breadth of application to many fields, e.g., control theory (engineering), stochastic parameter regression models (economics and finance), and Bayesian predictive methods (statistics). The view of the Kalman filter underlying its exposition here derives from Bayesian statistics [62]. In their application to modeling demands for aircraft spare parts, Kalman filters are a natural extension to Bayesian statistics in the sense that the probability distributions of the parameters underlying the demand process are updated with each new observation. As we will discuss at somewhat greater length below, the Kalman filter also has the important characteristic that its performance is fairly robust with respect to its underlying assumptions. In many applications, Kalman filters have been observed to perform better than expected when their assumptions are not met.

In the remainder of this section we introduce a simple Kalman filter model and examine its properties. Then we introduce a more general model and discuss an example of its application. We will then discuss a simplified version of this approach, weighted regression, that is easier to implement and show the results of our evaluations of a weighted regression approach to demand forecasting in terms of traditional statistical measures of forecasting error. Finally, we demonstrate its cost-effectiveness in estimating spares requirements in terms of aircraft availability and spares investment costs.

INTRODUCTION TO THE SIMPLE KALMAN FILTER

A Kalman filter model consists of two parts: a state equation and a measurement equation. The state equation describes the way the parameters vary over time. In the simple model that we are about to examine, the mean is the only parameter underlying the process that will be of interest. The state equation of the simplest Kalman filter model is given by

$$\beta_t = \beta_{t-1} + w_t \quad ,$$

where β_t is the mean of the process at time t , β_{t-1} is the value of the mean at time $t - 1$, and w_t is an error term that is $N(0, W_t)$, i.e., normally distributed with mean zero and variance W_t . This variance represents the sources of uncertainty that we have discussed in preceding paragraphs.

The measurement equation relates the mean to the observations of the process. The measurement equation of this process is given by

$$Y_t = \beta_t + v_t ,$$

where v_t is $N(0, V_t)$. β_0 has a normal prior distribution described by its mean and variance. In this application, the prior distribution of the mean will be derived from past observations of the process or of other, similar processes using an empirical Bayes approach, rather than by subjective judgment as is sometimes the case in Bayesian applications. In the following discussion the caret (^) is used to denote estimators and σ_t^2 will denote the specified variance of β_t at time t . Using this notation, the prior for β_0 is $N(\hat{\beta}_0, \sigma_0^2)$. The distribution of the mean will be updated with each observation of the process.

The idea is that we observe Y_t , a realization of the process whose mean is, momentarily, β_t . β_t is related to β_{t-1} but is not forced by assumption to be the same. This modeling procedure yields a natural approach to updating the distribution of β_t , a natural prediction method for future values of the process, and prediction error estimates that are more reasonable than those obtained from fixed parameter models. The error term, v_t , represents stochastic variance as in most statistical models. It affects one observation only; it is "noise" in the traditional sense. The w_t is an error term of a different sort. It is the difference between the underlying, nonstationary mean at two adjacent points in time. It is not transitory; i.e., it has a lasting effect on the process. It corresponds to a variation in the mean over time in the fashion of a random walk.

The computations of estimates and predictions using this model are straightforward. The estimation and prediction formulas are recursive; the estimate for β_t is computed from the estimate for β_{t-1} and from the current observation.

The estimate for β_t given $Y_t = Y_1, Y_2, \dots, Y_t$ is [63]:

$$\hat{\beta}_t = (1 - \alpha)\hat{\beta}_{t-1} + \alpha Y_t , \text{ where } \alpha = \frac{\sigma_{t-1}^2 + W_t}{V_t + \sigma_{t-1}^2 + W_t} .$$

This can be thought of as the result of mixing two estimates of known variances: $\hat{\beta}_t$ having variance $\sigma_{t-1}^2 + W_t$, and Y_t having variance V_t . This is nearly the same as simple exponential smoothing with a smoothing constant of α , except for the effect of the prior; however, the effect of the prior diminishes as additional observations of the process are made. In the long run, the predictions of the two ap-

proaches become indistinguishable. This is true only for certain special cases of the Kalman filter. More sophisticated Kalman filters cannot necessarily be equated to simple exponential smoothing.

The measure of uncertainty (the posterior variance) associated with the current state β_t is [63]:

$$\sigma_t^2 = V_t \alpha = V_t \frac{\sigma_{t-1}^2 + W_t}{V_t + \sigma_{t-1}^2 + W_t}.$$

Note that this is the harmonic mean of the two variances, $\sigma_{t-1}^2 + W_t$ and V_t . In Bayesian statistics, this is the usual posterior variance associated with a normal mean when the observation variance is known.

Several numerical values are required to support this method:

$$\hat{\beta}_0, \sigma_0^2, V_t, \text{ and } W_t.$$

Often, V_t and W_t are assumed to be unchanging over time. They describe, respectively, the likely magnitude of the measurement error and the period-to-period variability in the underlying mean. For some applications, it may be possible to estimate them quite accurately through data analysis; in other applications, V and W will be knobs that need to be tuned, with the results judged by methods that we will discuss below. In specifying the numeric values of W and V , we implicitly specify the ratio of stochastic variance to the sum of all of the other sources of uncertainty.

Estimating and Predicting the Updating Equations for This Special Case

We now examine the procedure for predicting k periods into the future and for attaching a measure of uncertainty to the prediction. Of course, k equal to 1 is the most common case.

Consider the problem of predicting for time $t + k$; note that

$$\beta_{t+k} = \beta_t + w_{t+1} + \dots + w_{t+k} \text{ and } Y_{t+k} = \beta_{t+k} + v_{t+k},$$

so the prediction is

$$E(Y_{t+k} | Y_t) = E(\beta_t + \sum_{j=1}^k w_{t+j} + v_{t+k} | Y_t) = \hat{\beta}_t.$$

Note that there is no deterministic direction in the variation of the underlying mean since the expected value of w_t is zero; thus the expected value of the process k periods in the future is the same as the current estimate of the mean, but the measure of uncertainty (predictive variance) grows with the length of the forecasting horizon as shown by

$$\text{var}(Y_{t+k} | Y_t) = V_t + \sigma_t^2 + kW_t$$

The three terms in $\text{var}(Y_{t+k} | Y_t)$ correspond, respectively, to measurement error in Y_{t+k} (stochastic variance), uncertainty about the current value of β (parameter variance), and the variability in β over the k time periods between t and $t + k$ (variance from all other sources). Intuitively, at least, this logic is superior to traditional methods of prediction. We know that if we are predicting far into the future we are less certain of our predictions. This logic reflects that uncertainty.

The expressions for $\hat{\beta}_t$ and σ_t^2 reflect that $\hat{\beta}_t$ is really most affected by the ratio of W to V , and that σ_t^2 is more affected by the magnitudes of V and W . The same is true for the predictions and the predictive variance: The prediction is mainly influenced by the ratio of W to V , the predictive variance more by the magnitudes of the variances.

Building Intuition About the Simple Kalman Filter Model

The Kalman filters that seem most appropriate for spare parts demand forecasting are somewhat more elaborate than this simple model. Nevertheless, building some intuition about this simple model may be useful in understanding the strengths and weaknesses of this forecasting approach. We will examine four archetypal data series and observe how the Kalman filter forecasts each one: step function, outlier, ramp, and oscillation. For each of these data series we show the Kalman filter's predictions one period ahead, in some cases with several different ratios of state variance to measurement variance. We will contrast these predictions to those from the traditional eight-quarter moving average. For all of these illustrations, the priors for the Kalman filters were set to a mean of 0 and a variance of 1. The effect of changing this prior is discussed below. As mentioned above, the predictions of a Kalman filter are primarily affected by the ratio of the state variance to the measurement variance. This ratio is included in the labels for the predictions used here; e.g., "Kalman 1:4" implies that the measurement variance is four times as large as the state variance.

Figure 4.1 shows the response of several Kalman filters and an eight-quarter moving average to a change in level (step function) in the data. Note that the higher the ratio of state variance to measurement variance, the more quickly the predictions adjust, because a high value of this ratio more heavily favors the observed data. Part of the art of Kalman filtering is adjusting this ratio to respond to real changes in the mean without making it so sensitive that it leaps around in response to noise in the data.

Figure 4.2 illustrates the response of a Kalman filter to a spike in the data. The prediction responds to the large data value but then dies down fairly quickly. The moving average does not react as much but the reaction persists for a longer period.

Figure 4.3 shows the response of the Kalman filter to a ramp in the process. Note that the Kalman filter's predictions are closer to the actual data than the predictions of the eight-quarter moving average in all but one period. The Kalman filter tracks the ramp better than the moving average and then responds quickly to the return to zero.

Figure 4.4 illustrates the response of the Kalman filter to an oscillating process. This is the kind of process that might result from annual cycles in the data. This example illustrates when a simple Kalman

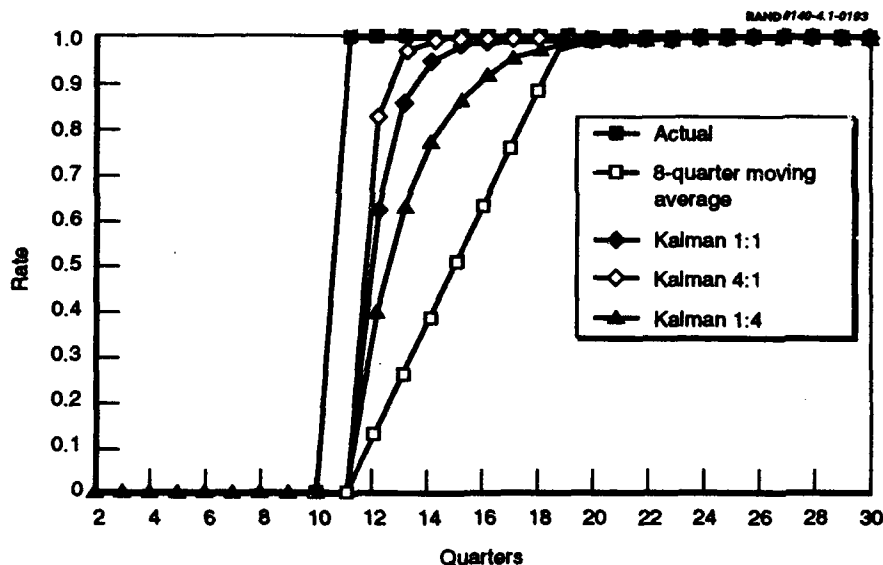


Figure 4.1—A Step Function

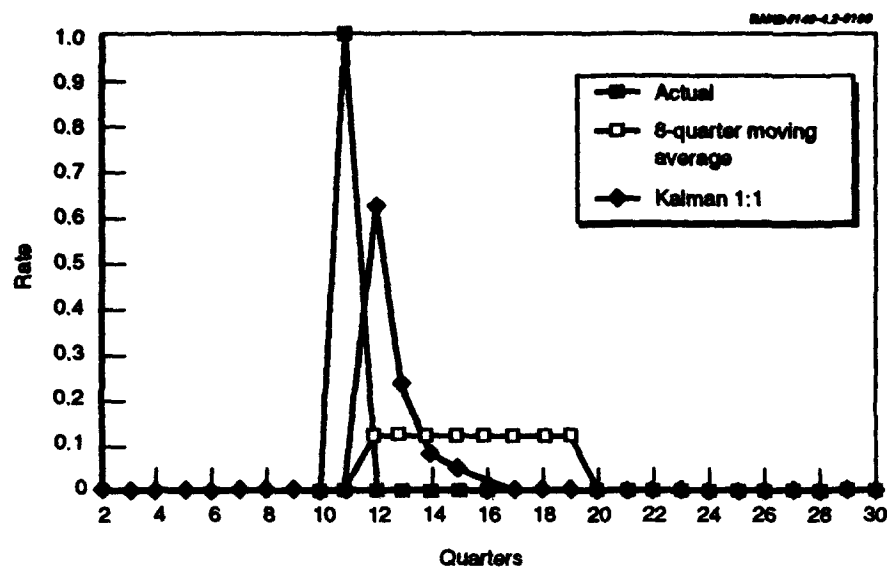


Figure 4.2—An Outlier

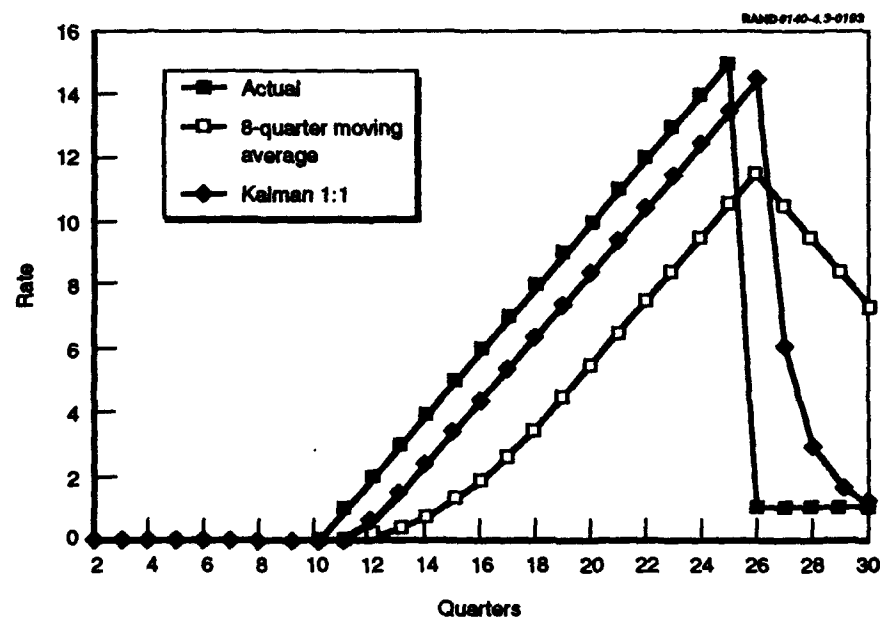


Figure 4.3—A Ramp

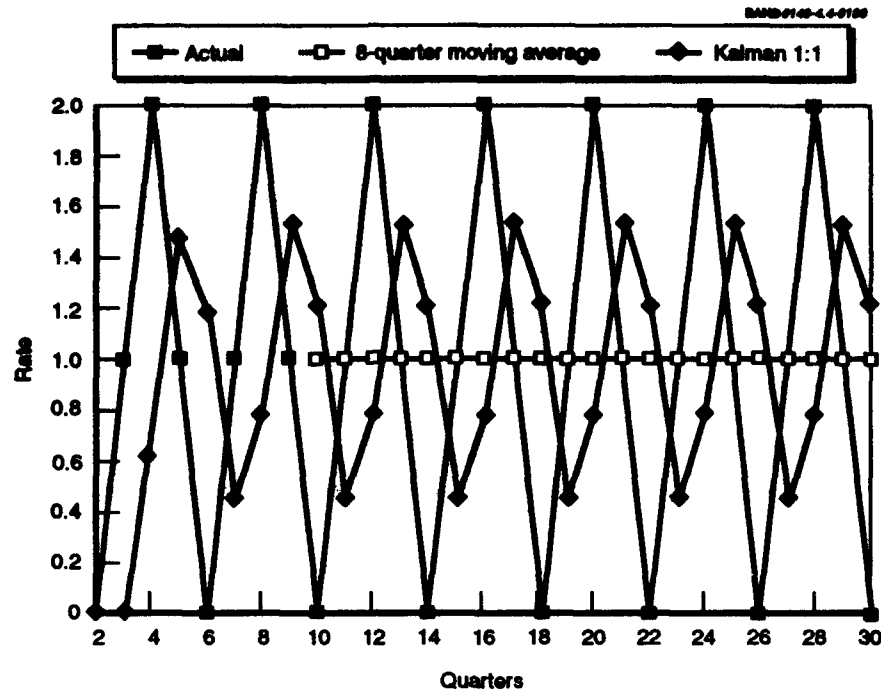


Figure 4.4—An Oscillating Process

filter might be inappropriate. More elaborate Kalman filters have been designed to deal with cyclic behavior. If a particular Kalman filter model forecasts poorly, there are many extensions and enhancements available to treat most common forecasting problems.

Figure 4.5 addresses the effect of the prior on the Kalman filter's predictions. This graph changes the prior mean to 5.0 and leaves all other values unchanged. The effect of the prior has disappeared by the eighth period. Since we are comparing the forecasts to those from an eight-quarter moving average that doesn't start producing forecasts until period nine, the prior has no effect here. In general, the prior could have an effect after eight periods if the state variance and the prior variance were very small or if the prior mean were very large. In later applications in this report we will use noninformative priors and will be careful to avoid highly influential (informative) priors.

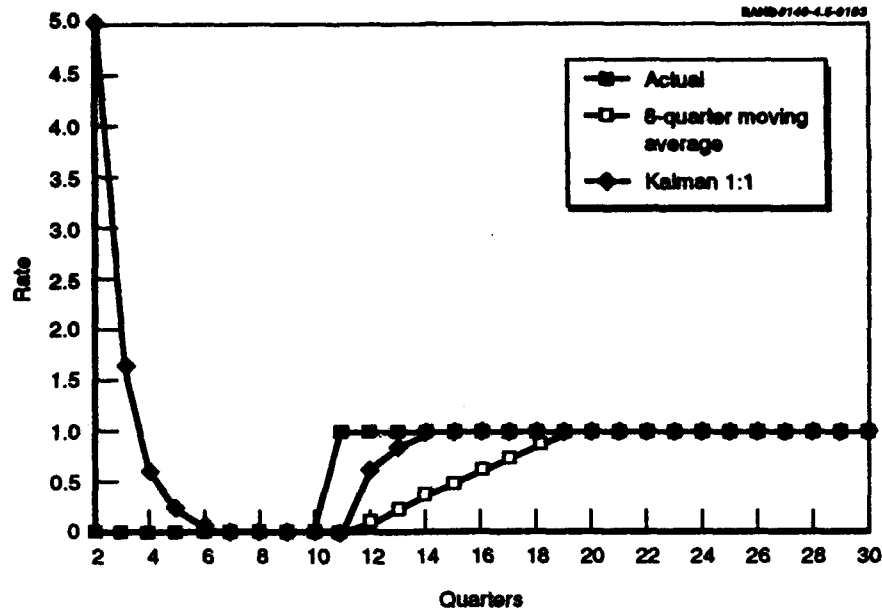


Figure 4.5—The Effect of the Prior on the Forecasts

INTRODUCTION TO THE GENERAL KALMAN FILTER MODEL

A general Kalman filter model allows us to handle more realistic measurement equations and state equations. Explanatory variables may be added to the measurement equation and trends or cycles may be added to the state equation. The state equation can allow different rates of change for different parameters of the model. These features make the general model especially flexible and powerful in modeling more complicated processes.

Development of the General Kalman Filter Model

The Kalman filter model can be generalized to vector-valued parameters and multivariate observations. The formulas are quite similar, except that vectors and matrices replace the scalar values. The state equation of the general model is given by

$$\beta_t = A_t \cdot \beta_{t-1} + w_t ,$$

and the measurement equation by

$$Y_t = X_t \cdot \beta_t + v_t ,$$

where w_t is distributed $N(0, W_t)$ and v_t is distributed $N(0, V_t)$, where the N denotes the normal distribution. β_0 has a normal prior described by its mean and variance. β_t is a column vector, A_t is a square matrix, and X_t is a matrix. Note also that X_t need not be square and Y_t and β_t need not be of the same length. Indeed Y_t is often a scalar. W_t and V_t are now covariance matrices rather than scalar variances. We use this "normal-normal" model for computational convenience. Below we discuss the use of transformations to fit the discrete nature of the data better.

This more general model will be used in the demand forecasting example given below and will be discussed as it applies to demand forecasting in following sections.

Building Intuition About the General Model

It may be helpful to examine an application of this general model to a specific example. The following example uses data on a component of the F-111 aircraft, an airspeed indicator.

In this application, the number of observed demands in a time period (quarter) is modeled as

$$D_t = \beta_{0t} + \beta_{1t}f_t + v_t ,$$

given β_{0t} and β_{1t} , where f_t denotes hundreds of hours flown by aircraft in which the airspeed indicator is installed; i.e., $f_t = 20$ means that the airspeed indicator experienced 2,000 hours of use. This is the measurement equation. The error term v_t is normal with variance V_t to be specified. The quarter-to-quarter evolution of β_t will be modeled by the state equation

$$\beta_{t+1} = \beta_t + w_t ,$$

where w_t is a bi-variate normal with zero covariance and variances W_{1t} and W_{2t} to be specified. β_t is the vector containing β_{0t} and β_{1t} . Thus, in the notation of the previous example,

$$\beta_t = (\beta_{0t}, \beta_{1t}) \text{ treated as a column vector,}$$

$$F_t = [1, f_t] \text{ treated as a } 1 \times 2 \text{ matrix.}$$

Some things need to be specified to start up the Kalman filter machinery. In the current problem, these *eight numbers* are

- The mean and covariance matrix for β_0 : two means, two variances, and a covariance,
- The two state variances W_{1t} and W_{2t} (the state covariance is specified to be zero here), and
- The measurement variance V_t .

This case differs from the simple example in that, in the former case, the state and the measurement were on the same scale. In this case, the state is described by a vector of regression coefficients, and the measurement is an observed number of demands. They are connected by F_t . One implication of this is that the state and measurement variances are on different scales, which makes them somewhat harder to compare. The state propagates by an identity matrix; i.e., the only change from one period to the next is stochastic.

In this discussion, we will use a somewhat ad hoc method of setting the starting values for the filter. For the prior values we will just regress the first eight quarters of demands on the flying hours and use the parameter estimates from this regression for our prior values for β_0 . We will use the estimated covariance matrix from this regression for σ_0^2 . Setting the state and measurement variances will be a two-stage procedure. First, we will divide the variance of the first eight quarters of demand and explore alternative allocations of that variance between the state and measurement equations. Second, in the state equation we will divide the variance between β_{0t} and β_{1t} equally. Half of the state variance will be assigned to β_{0t} ; the other half will be divided by the mean of the first eight quarters' flying hours and assigned to β_{1t} . This division will make the contribution from the variance of β_{1t} (after it has been multiplied by flying hours) to the overall variance roughly equal to the contribution from β_{0t} .

A full Bayesian treatment of this problem would require much more careful thinking about the prior values for each part. Since it is ultimately impractical for us to engage in a careful specification of the prior for each of the parts in the Air Force inventory system, we must use simpler rules. We can evaluate different rules for specifying the priors by observing the quality of their predictions over a wide range of parts. This is an example of a pragmatic application of the Bayesian paradigm. In a discussion below, we will use a more easily implemented approximation to the Kalman filter, the *weighted regression* forecaster.

Figures 4.6 and 4.7 illuminate the effects of the variances on the predictions of this type of model. Figure 4.6 shows the results of making the state and measurement variances roughly equal. Note that the predictions generally increase with the observed data until the middle of the sample period and then drop down. Generally, the pattern is that predictions lag the demands by one period; i.e., if there is a drop in demands, the predictions typically drop in the following period. This is not necessarily the case, however; demands increased slightly in period 26 but the prediction decreased in period 27 owing to a change in flying hours.

Figure 4.7 illustrates what happens if the state variance is set to zero, corresponding to fitting a regression to all the data up to time t and then using the regression to predict for time $t + 1$. Notice that in the middle of the sample period the predictions underpredict demand. This happens because this model is still giving full weight in its predictions to the early periods, but the world has evolved away from the demand rate that described this early period.

Values of state and measurement variances that are between these two cases would provide intermediate levels of smoothing. Note that making the state variance large will not produce a flat horizontal line;

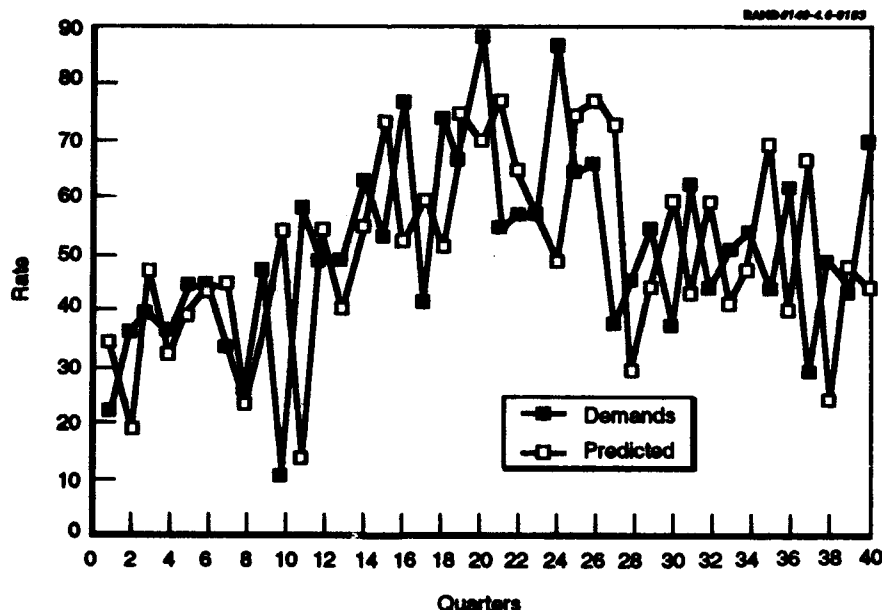


Figure 4.6—State Variance and Measurement Variance Equal

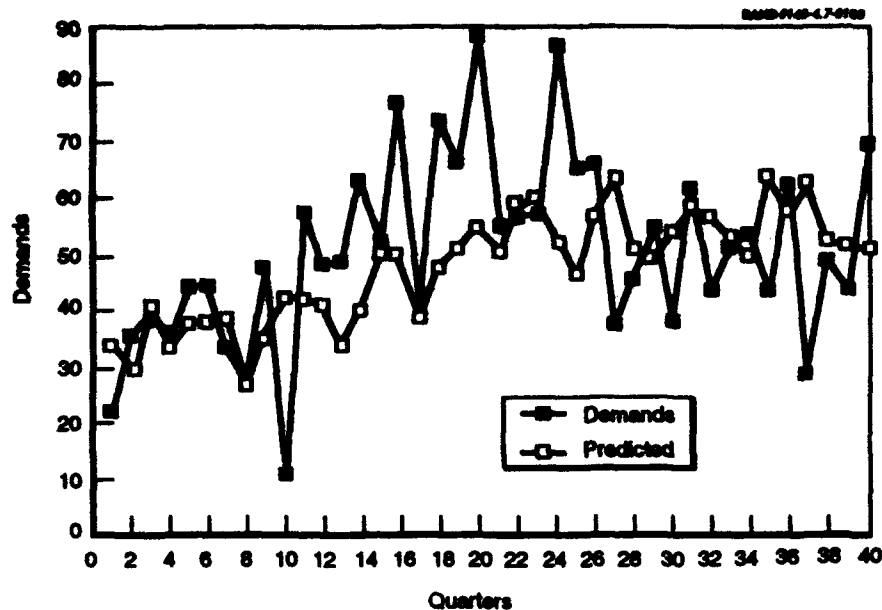


Figure 4.7—State Variance Set to Zero

i.e., it smooths the underlying regression parameters, not the demands. The effects of the variation in flying hours will still be seen in this graph.

Robustness of Kalman Filters

Why is it advantageous to use time-varying coefficients instead of a bigger variance on the error term v_t ? Suppose some function relates expected demands to flying hours.

- If the function is linear and unchanging in time, then the Kalman filter has to do worse than a model that assumes fixed coefficients, because the Kalman filter makes less use of older observations when they are just as good as recent ones. An example would be using the sample mean as opposed to exponential smoothing to estimate a stationary population mean, a case in which exponential smoothing is inefficient.
- If the function doesn't vary in time and is nonlinear, the Kalman filter will be better than a fixed parameter model if the flying hours are positively autocorrelated in time, because at any given time the Kalman filter is linearly approximating the correct part of the

curve—that is, the part of the curve associated with the current flying hours—because of the presumed autocorrelation of flying hours.

- If, on the other hand, the curve really is linear but the slope and intercept vary over time (for whatever reason, such as omitted variables), the Kalman filter works better than the fixed parameter model if the slope and intercept are positively autocorrelated in time: Old data tell you less about where the line is than do current data. Again, if the mean is a random walk, the exponential smoother will beat the sample mean in estimating the current value if the variance of the random walk (W_t) is large enough.

In general, we may observe both nonlinearity in the relationship between flying hours and expected demands and a changing relationship over time. The Kalman filter, in effect, smooths in both time and in the regressors, thus achieving greater adaptability in the face of changing parameters and flying hour programs. If the required autocorrelations are present, the Kalman filter should do better. This may help explain the observed (but somewhat mysterious) robustness of the Kalman filter.

Relationship to Other Prediction Methods

One useful characteristic of Kalman filter models is the way they generalize other well known time series and forecasting methods. Unfortunately, this has caused some confusion in the literature. It is not correct to claim that Kalman filtering is “just exponential smoothing” as is sometimes heard. Exponential smoothing is a special case of Kalman filtering but Kalman filtering incorporates a much richer collection of models. In Appendix A, we discuss some well known statistical models and how they can be written as Kalman filter models.

5. MORE FLEXIBLE DEMAND AND NRTS FORECASTS

In this section we will apply two alternative forecasting methods to the demands for a sample of aircraft spare parts. Both of these methods are motivated by a need for greater flexibility to cope with the nonstationarity of the demand process. The two basic types of prediction methods we consider here are (1) weighted calculation of demand rates, and (2) weighted regression demand forecasters. We also examined the performance of Kalman filter regression forecasters in this research but concluded that the weighted regression methods performed at least as well and were more easily implemented. The weighted regression forecaster can be thought of as a special limiting case of the Kalman filter regression forecaster that incorporates a noninformative prior; therefore, it has the special advantage in an application as large as the Air Force's recoverable spares inventory system of not requiring the specification of the priors needed for the Kalman filter regression forecaster [63].

The quality of predictions is evaluated here in two different ways, root mean squared error (RMSE) and mean absolute deviation (MAD). In subsequent discussion, we evaluate the predictions in actual application to spares requirements computations with a capability assessment model, Dyna-METRIC Version 6. Root mean squared error is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\text{back-transformed predicted value}_i - D_i)^2}{n}},$$

where D_i is the demand in period i and the back-transformed predicted value is typically the prediction from the model. In a few of the models considered here, the forecast variable is in the square root scale, so back-transforming (in this case squaring) is required to put the predictions in the appropriate scale.

Mean absolute deviation is defined as

$$MAD = \frac{\sum_{i=1}^n |\text{back-transformed predicted value}_i - D_i|}{n},$$

where the $| |$ symbol stands for the absolute value. This summary of

the quality of the predictions weights larger predictive errors less severely than RMSE does. Whether these or any other measures of predictive quality are appropriate depends on the application, but these measures will do for illustrative purposes.

Throughout this section we will use these measures of predictive quality to compare alternative forecasting methods with the forecasting method based on the eight-quarter moving average now used by the Air Force.

Since our motivation for exploring alternative demand modeling and forecasting techniques is to improve the estimation of requirements for aircraft recoverable spares and depot-level repair, we are interested in the performance of alternative forecasting techniques over the time horizons applicable to the requirements estimation problem. For depot-level repair, as a practical matter, the forecasting horizon is only two quarters. Although repair planning is also done over longer planning horizons, the Air Logistics Center has considerable quarter-to-quarter flexibility in adjusting repair quantities. For spares procurement purposes, however, planning horizons are typically about three years long because of the source selection and manufacturing processes involved. Exploration of methods that might help alleviate those constraints are beyond the scope of this work. Therefore, we will deal with these longer horizons in the requirements forecasting problem and will examine the performance of alternative forecasting methods over planning horizons as long as 13 quarters.

THE PARTS SAMPLE

The sample of data used in these examples consists of BP15 recoverable aircraft spares from D041 with one or more demands in the previous eight quarters and a program select code of one (i.e., the demands for these parts are assumed to be driven by flying hours). The time period of the data runs from March 1980 to March 1989 by quarters. The sample was further restricted to parts that had incurred more than four demands systemwide in at least one of the 40 quarters. There is little reason to expect that the techniques we consider here can improve demand predictions for parts that never generated more than four demands in a quarter over 40 quarters. We also restricted the sample to parts that had valid demands (zero is a valid demand; missing is not) in at least 12 quarters. Although the prediction techniques considered here need at most eight quarters of data to get started, we need a few quarters of data with which to evaluate the quality of their predictions.

Missing values and zero demands are issues in these data. We use an operational definition: If there are flying hours in the record for the part and zero or missing demands, then we infer that the part had zero demands. If the part has zero demands and the flying hours are missing, then the demands are missing.

These restrictions left us with 14,007 parts in our universe. We stratified these parts into three groups. Group 1 consisted of parts with an average demand of more than 15 per quarter in the quarters where the demands were not missing from the database (4,215 parts). Groups 2 and 3 both had less than 15 demands per quarter on average. Group 2 consisted of parts that cost less than \$2,500 (4,966 parts). Group 3 consisted of parts that cost more than \$2,500 (4,826 parts). From each of these three groups we selected a random sample of 200 parts for the evaluations.

METHODS AND RESULTS

In the discussion that follows, we explore the performance of weighted calculation of demand rates and weighted regression forecasters.

Weighted Calculation of Demand Rates

A simple compromise between the more sophisticated forecasting methods considered in this section and the Air Force's current eight-quarter moving average forecasts is weighted demand rates. The current system calculates the demand rate as

$$\text{Demand rate}_{t+k} = \frac{\sum_{i=t-7}^t D_i}{\sum_{i=t-7}^t fh_i},$$

where D_i denotes the i th quarter's demands, fh_i the flying hours in the i th quarter, and k the number of quarters into the future. Forecasts are then made by simply multiplying this rate times the future program:

$$\text{Predicted demands}_{t+k} = \text{demand rate}_{t+k} fh_{t+k}.$$

A simple way to make this model more responsive is to weight recent observations more heavily than older observations:

$$\text{Demand rate}_{t+k} = \frac{\sum_{i=t-7}^t w_i D_i}{\sum_{i=t-7}^t w_i f h_i}$$

A Kalman filtering approach suggests weights of the form $w_i = \alpha^{t-i}$. We tried values of α of 0.25, 0.5, and 0.75; the results for $\alpha = 0.25$ were aberrant and are not included here. We also tried a four-quarter moving average without weights as another ad hoc way to gain some responsiveness in the forecasts. Figures 5.1 and 5.2 reflect the root mean squared error and mean absolute deviation, respectively, over various time horizons up to 13 quarters in length, for the four approaches. These are calculated from the sums of the predictive quality measures over all the parts in Group 1. We tried these methods for Groups 2 and 3 as well, but they did not perform better than the current approach, possibly because of the failure of the normality assumption or insufficient information content in low-demand data. These results show that weighted demand rate forecasters perform better than does the current approach. The fact that the four-quarter moving average also does better than the current approach suggests that the current approach is not sufficiently responsive to non-

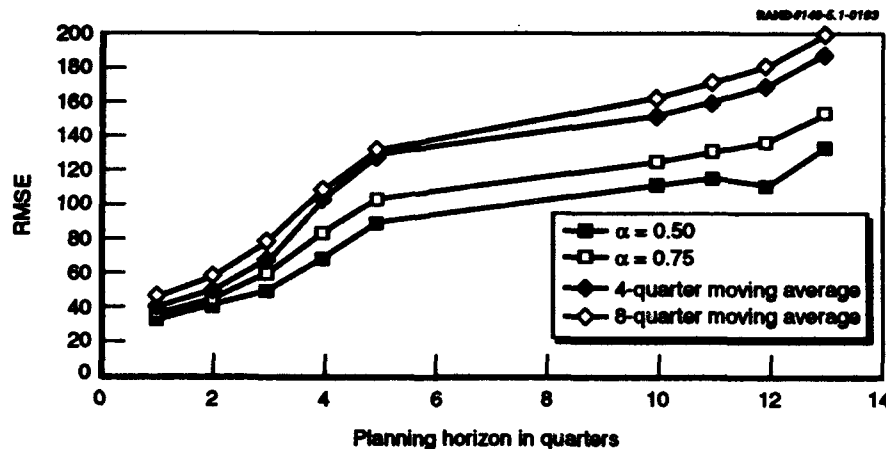


Figure 5.1—Root Mean Squared Error of Weighted Demand Rate Forecasters for High-Demand Parts

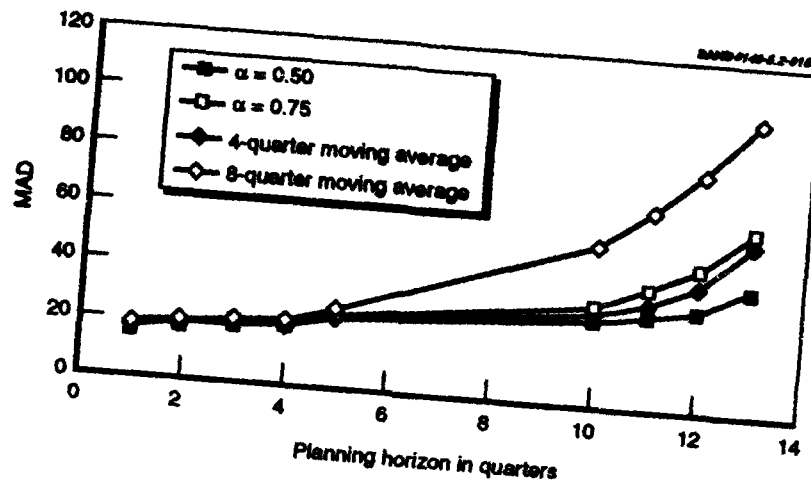


Figure 5.2—Mean Absolute Deviation of Weighted Demand Rate Forecasters

stationarity. The better weighted demand forecaster used a weighting factor of 0.5. It is clearly superior to the eight-quarter moving average, and its superiority increases with the length of the forecasting horizon.

The difference between the RMSE and the MAD is due to the differing weights they place on large errors. The RMSE penalizes large errors more harshly than MAD.

Weighted Regression Demand Forecasters

In this section we consider forecasts based on moving regressions. For a moving eight-quarter window, we fit a regression of the form $D_t = \beta_0 + \beta_1 t$, then use the model to predict the demands for the following quarter. We do this with various exponential weights and with weights of unity with both eight-quarter and four-quarter windows.

The weighted regressions in Figures 5.3 and 5.4 use exponentially declining weights in an eight-quarter window. That is, eight quarters of data are used with the most recent quarter given a weight of α ($0 < \alpha \leq 1$), the second most recent given a weight of α^2 , and so on. Note that large improvements are achieved by moving from weighted demand forecasters to weighted regression forecasters for any value

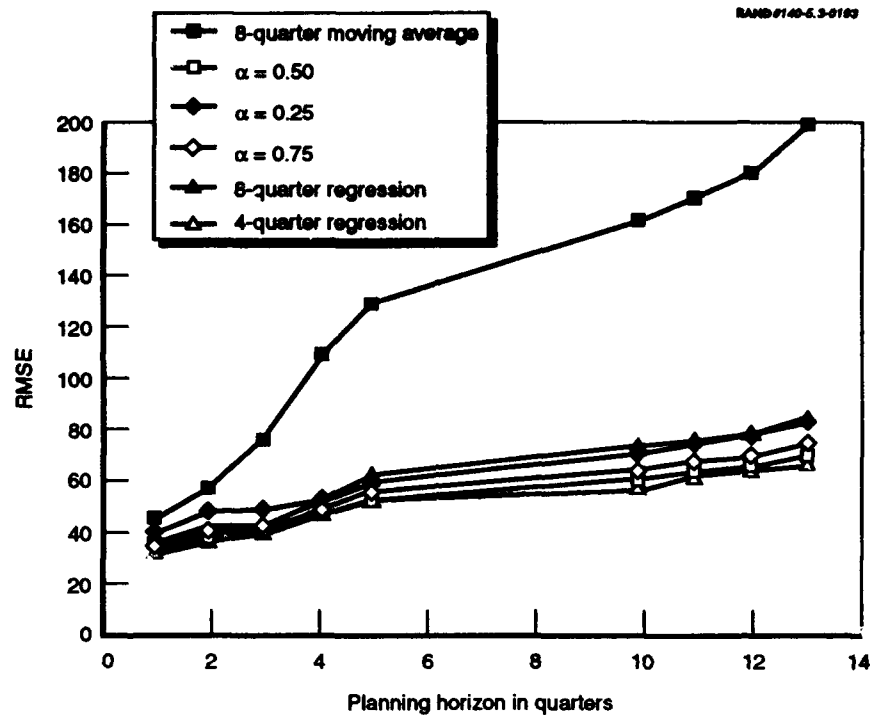


Figure 5.3—Root Mean Squared Error of Weighted Regression Forecasters

of α . However, there are additional gains to be made by adjusting the value of α for both approaches.

THE COMPUTATIONAL ALGEBRA

The weighted regression forecaster we recommend is specified in the following algebra. Suppose we observe the most recent eight quarters of demands and past item flying hours in the sequence d_1, d_2, \dots, d_8 , and fh_1, fh_2, \dots, fh_8 , respectively. We will use the weighting factors w_1, w_2, \dots , and w_8 to assign greater weight to the more recent quarters by setting the $\{w_i\}$ equal to 0.75^{8-i} , $i = 1, 2, \dots, 8$. Thus the weights will be

$$w_1 = 0.75^7 = 0.1335, w_2 = 0.75^6 = 0.1780, w_3 = 0.2373, \\ w_4 = 0.3164, w_5 = 0.4219, w_6 = 0.5625, w_7 = 0.75^1 = 0.75, \text{ and} \\ w_8 = 0.75^0 = 1.0. \text{ The sum of } \{w_i\} = 3.5995.$$

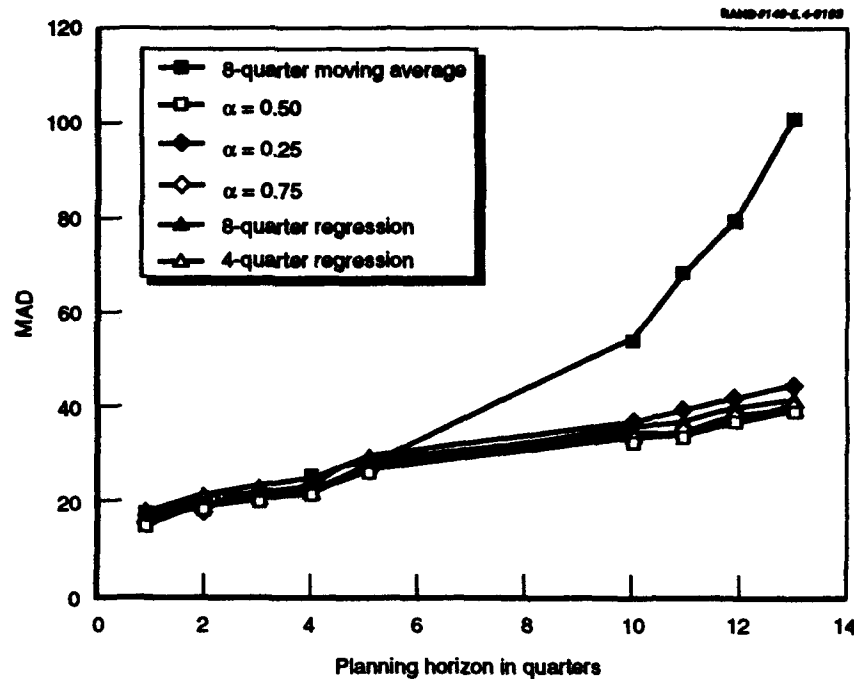


Figure 5.4—Mean Absolute Deviation of Weighted Regression Forecasters

Let the notation Σ_i be defined to mean the sum taken over the eight quarters. The weighted mean demand, D^* , is given by

$$D^* = \frac{\sum_i w_i d_i}{\sum_i w_i} ,$$

and the weighted mean flying hours (item program), fh^* , by

$$fh^* = \frac{\sum_i w_i fh_i}{\sum_i w_i} .$$

We define

$$\Sigma_{xy} = \left(\sum_i w_i \right) \left(\sum_i w_i d_i fh_i \right) - \left(\sum_i w_i d_i \right) \left(\sum_i w_i fh_i \right) ,$$

and

$$\Sigma_{xx} = \sum_i w_i \left[\sum_i w_i (fh_i)^2 \right] - \left(\sum_i w_i fh_i \right)^2 .$$

Then

$$\beta_1 = \sum_{xy} / \sum_{xx} \text{ and } \beta_0 = D^* - \beta_1 fh^*$$

If β_1 is negative, revert to the eight-quarter moving average.

NRTS FORECASTING

NRTS forecasting is especially important in depot repair planning. NRTS actions are one subset of base-level demands, the other two being repairs and condemnations. Over longer planning horizons, the current methods are subject to large errors. We explored alternatives to the current method and describe the results here.

An improved method of forecasting NRTS actions results from combining the improved demand forecasting method with an improved method of forecasting the NRTS rate. Again, the method is recommended for high-demand items. The forecast of the NRTS rate is done by weighting more recent observations of the NRTS rate more heavily than older observations using a simple exponential smoothing technique and a weighting factor of 0.75. In the notation just introduced,

$$\text{NRTS rate} = \sum_i w_i \text{NRTS}_i / \sum_i w_i d_i,$$

where NRTS_i is the number of NRTS actions in quarter i .

This NRTS rate estimator, combined with the improved demand forecasting technique already discussed, delivers the performance reflected in Figures 5.5 and 5.6 for high-demand items. The alternative method dominates the current method in root mean squared error over all planning horizons examined. In mean absolute deviation, though, the current method does slightly better for shorter planning horizons. The alternative method is our choice for any planning horizon.

CONCLUSIONS

The compelling message of Figures 5.3 and 5.4 is that any one of these demand forecasting methods clearly dominates the eight-quarter moving average and their superiority increases with the length of the forecasting horizon. The question remains, then, of which method to choose. For two reasons, our choice is the weighted regression forecaster with an α of 0.75: (a) It scored best in its average ranking

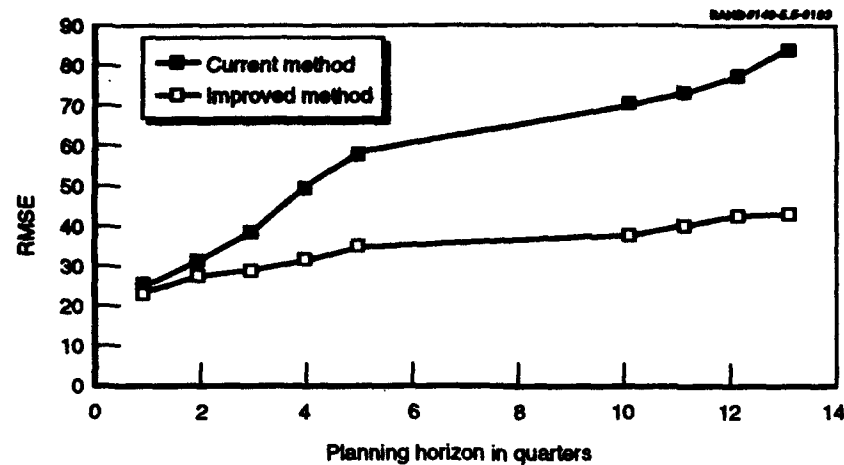


Figure 5.5—Root Mean Squared Error of Improved NRTS Forecaster

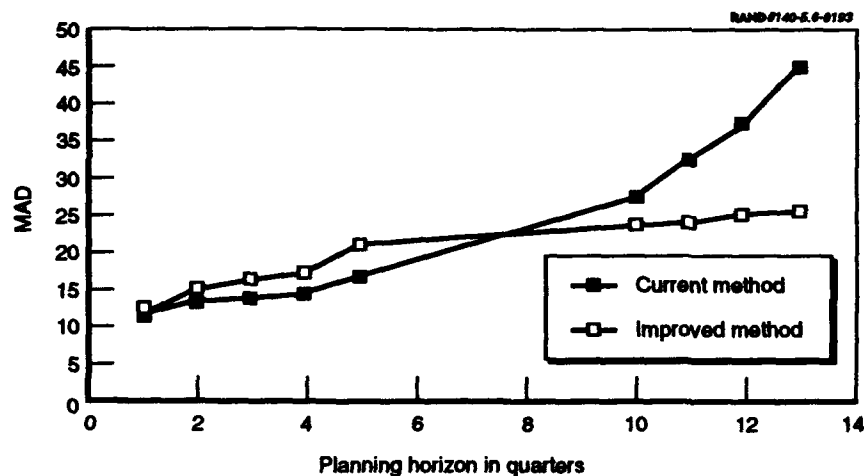


Figure 5.6—Mean Absolute Deviation of Improved NRTS Forecaster

among RMSE and MAD scores relative to the other methods, and (b) it is more intuitively appealing than methods that effectively use any less of the available data. For example, the four-quarter moving average did quite well in terms of mean absolute deviation but depends on four quarterly observations to estimate two parameters, hardly an appealing idea. It is also appealing that the weighted regression forecaster with an α of 0.75 does not have such a steep slope in its

weighting of the eight quarters of observations as do weighted regression methods with smaller values of α .

It is important to understand that the apparent superiority of the weighted regression demand forecasters over the weighted demand rate forecasters partially derives from the inclusion of a non-zero intercept in the model of demands as a function of flying hours. This feature is important; it departs from an assumption of strict proportionality between demands and flying hours. The weighted regression demand forecasting procedure explicitly estimates the intercept.

By both of our evaluative criteria, the $\alpha = 0.75$ weighted regression demand forecaster does fairly well for Group 1. Therefore, for items in Group 1, we recommend it as the default demand forecaster for spares requirements computations. In any event, the use of the eight-quarter moving average should be discontinued for high-demand items.

For NRTS forecasting, the improved NRTS forecaster incorporates the improved demand forecaster and adds an improved NRTS rate forecaster that also uses a 0.75 weighting factor. It dominates the current method in root mean squared error over all planning horizons examined and dominates the current method in mean absolute deviation over long planning horizons. We recommend its use as the default NRTS forecaster for items in Group 1. We propose no change for Groups 2 or 3.

6. VARIANCE ESTIMATION

Estimating the variance of demands for aircraft recoverable spare parts is difficult for two reasons: (a) The statistical problem of variance estimation is, in itself, difficult because estimators of variances typically have very high error variances themselves, and (b) because our model of the demand process is so imperfect, the effects of many factors on the observed variability in the process are lumped into our estimate of the variance, as we discussed in Section 4. In this section, we discuss some characteristics of the VTMR estimator introduced in Section 3, suggest an improvement to variance estimation that corrects a logical flaw in the current system's specification of variance and, finally, mention a few approaches to variance estimation that did not work well.

SOME OBSERVATIONS ABOUT THE VTMR ESTIMATOR

To understand better the role of nonstationarity and data partitioning in estimating the underlying VTMR, we explored some of the characteristics of the VTMR estimator in both stationary and nonstationary processes: (a) The variance of the estimator increases with the coarseness of the partitioning and the true VTMR even when the process is stationary, and (b) very serious overestimation of the true VTMR (i.e., the VTMR of demands around the mean, stationary or nonstationary) occurs in the presence of nonstationarity, and the overestimation is an increasing function of the demand rate. These findings are fundamentally important to any use of the VTMR estimator in spares and repair requirements computations. They are consistent with (a) the association of larger values of observed VTMRs with high-demand items, (b) the substantial differences between the distributions of observed VTMRs discussed in the context of the Bitburg data analysis and those discussed by Crawford [2], and (c) the wildness of the distribution of observed VTMRs around the power function used by AFMC to specify VTMRs to describe the probability distribution of the number of items of each type in resupply. Forecast variance is discussed below.

Consider a simple Poisson demand process with a mean of 10 per quarter. Given this stationary process and eight quarters of observations, the VTMR estimator has a standard deviation of 0.53. Even with this very well behaved process, we can expect to observe a VTMR of about 1.9 or more about five times out of 100. Thus, in ob-

serving a substantial number of parts demand histories, one should not be surprised to see some large VTMRs, even if the underlying processes are simple Poisson. It may be incorrect to conclude that some of the processes have large underlying VTMRs. The risk here is that we may tend to confuse the large variability in the sampling distribution of the VTMR estimator with variability in the underlying process.

The same risk applies to inferences in the other direction. We may observe a low VTMR for any individual part when, in fact, the VTMR underlying its demand process is substantially greater than the value of the estimator. One approach to reducing our vulnerability to error in estimating the variability underlying a demand process for an individual part is to pool our estimates of variance across parts. In the current system, this pooling is accomplished through the application of the power function discussed previously. It may seem unsatisfying, but it does eliminate much of the instability in variance estimation across time that would occur if variances were estimated part by part. Part-specific variance estimation would also induce much greater volatility in numerical values of the AFAO over time, in turn inducing procurement actions in response to momentary changes in the estimated variances of specific parts.

Effects of Bias, Data Partitioning, and Nonstationarity on the VTMR Estimator

We describe briefly in the paragraphs that follow the bias function of the VTMR estimator and its mean and standard deviation as a function of data partitioning, mean demand rate, underlying VTMR, and two specific examples of nonstationarity in demand. The results are important in their implications for spares and repair requirements estimation as well as in explaining some of the observations of earlier researchers in this area. They also serve to underscore the difficulty of the VTMR estimation problem.

In Figures 6.1 through 6.3, we hold the total time period constant and subdivide it into periods of various lengths. The mean we refer to is the mean number of demands per time period.

Figure 6.1 reflects the VTMR estimator's bias function with demand processes with underlying VTMRs of 1.0, 2.0, and 3.0. With a stationary process, the expected value of the VTMR estimator is unaffected by the partitioning of the data. The bias is very modest above 50 or so total demands in the time period. Since the high-demand parts tend to drive the performance of the inventory system, the effects of the bias function are probably not important in shaping sys-

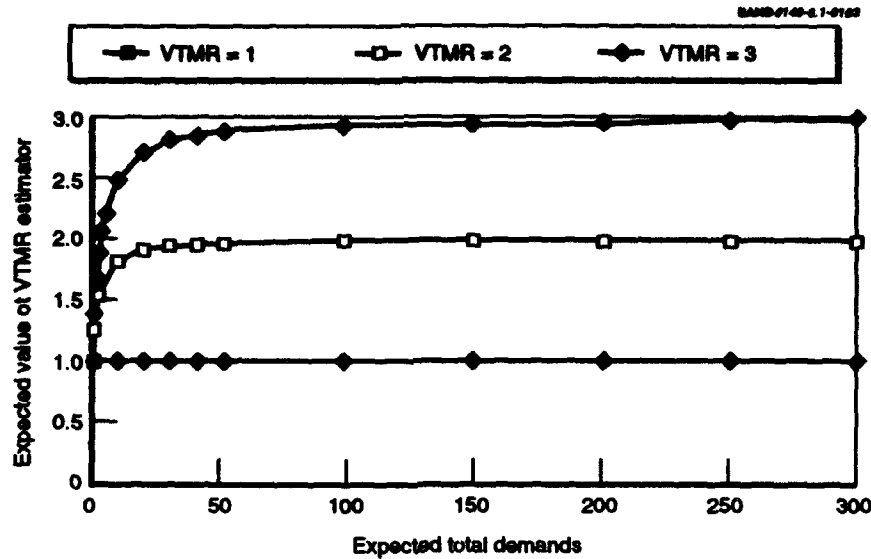


Figure 6.1—Effect of the Estimator's Bias Function

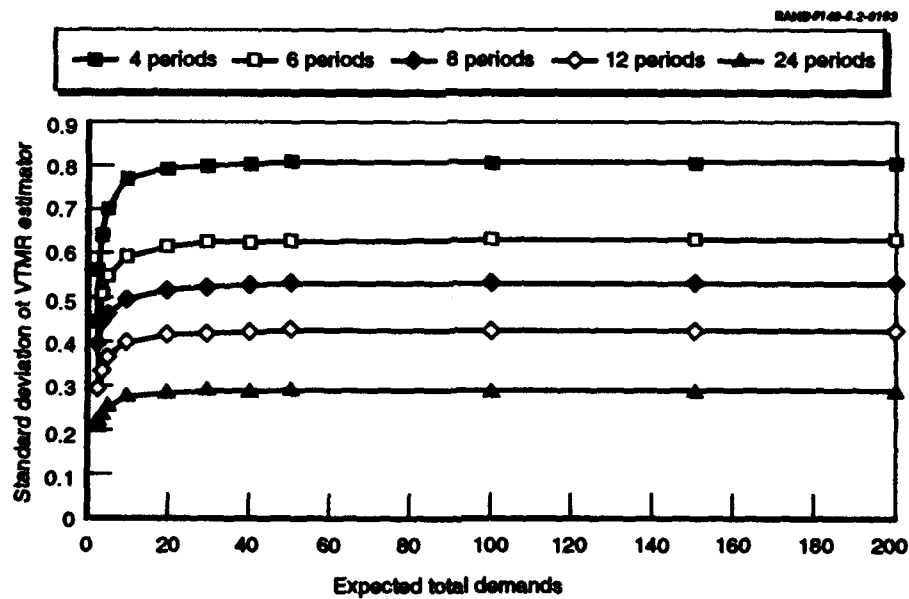


Figure 6.2—Effect of Data Partitioning on the Standard Deviation of the VTMR Estimator

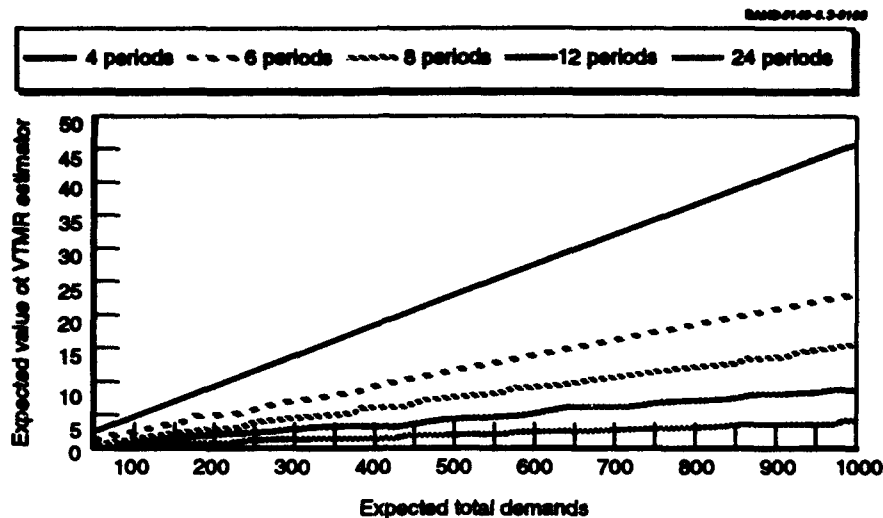


Figure 6.3—Effect of Data Partitioning on the Expected Value of the VTMR Estimator

tem performance. At the systemwide level used by AFMC to compute spares requirements, the bias function does not play nearly as important a role in VTMR estimation as nonstationarity, as we will show. The effects of nonstationarity can entirely swamp the bias by inducing estimation error in the opposite direction, i.e., overestimation.

Figure 6.2 illustrates the effect of data partitioning on the standard deviation of the VTMR estimator with a stationary Poisson process. Each curve represents a different partitioning of the data as explained in the legend. The coarser the partitioning, the higher the standard deviation of the estimator. Note that expected total demands has very little effect. The slightly smaller values associated with expected total demands of five are probably due to the default value of 1.0 being assigned to the estimator for the random realizations in which no demand occurred. *In this particular exposition, partitioning alone almost triples the standard deviation of the estimator.* The current system operates with quarterly data; therefore, it is vulnerable to the phenomenon shown here. A finer partitioning of demand data could decrease the standard deviation of the VTMR estimator.

Nonstationarity: A Ramp Process

We next examine a nonstationary process whose mean changes linearly during a two-year period from 0.5 to 1.5 times the overall mean. With a partitioning of the data into 24 equal periods (roughly corresponding to months) and expected total demands of 50, for example, the mean demand rate per period is $50/24 = 2.0833$. In the case of this ramp process, the expected demands per period are 1.0417, 1.1322, 1.2228, . . . , 3.125. Figure 6.3 shows the effects of partitioning and expected total demands on the expected value of the VTMR estimator in the case where variation around the changing mean follows a Poisson process. Both treatments have remarkable effects.

We direct the reader's attention to the middle curve in Figure 6.3, i.e., the curve representing a partitioning of the data into eight periods, as is the case in the current system. The length of the periods is roughly a fiscal quarter. Note that the expected value of the VTMR estimator increases from about 1 to 14 as a function of the total demands in the two-year period. Although the increase in the mean from 0.5 to 1.5 times its average value is rather steep, it does serve to underscore the dramatic effect that nonstationarity can have on the observed VTMR.

This is only one of the two important effects reflected in this illustration. Note that, for an item with 1,000 total demands in the two-year period (an average of 125 per "quarter," an unusually high-demand item), partitioning of the data has a dramatic effect on the expected value of the VTMR estimator. Moving from eight periods to 24 periods in the two years of demands reduces the expected value by about two-thirds. Even as we move toward the left of Figure 6.3 to smaller values of total demands, finer partitioning of the data still has major effects for items whose demand rates are more typical of those in the inventory system. Of course, if we knew that the underlying mean was a ramp process, we wouldn't estimate the VTMR in the traditional way. The point is to illustrate that the direction of the error is typically positive.

The combination of these two effects, coarseness of partitioning and nonstationarity, can result in some very large values of the VTMR estimator when the demand process is really quite well behaved around a nonstationary mean. There is little question that we are over-estimating VTMRs in the current system when we view those VTMRs as representing the stochastic variability in the demand process. The problem is that in estimating spares requirements, stochastic variability is not the only uncertainty we face in the forecasting problem. We also face all of the other sources of uncertainty associated with es-

timating the mean demand and variability of the process in the future, i.e., at the end of the procurement lead time.

Our principal observations about the VTMR estimator are:

- The variance of the VTMR estimator, r_j , increases with the coarseness of the partitioning of the observed data even when the process is stationary.
- Nonstationarity acts to induce overestimation of the VTMR as a measure of variation about the changing mean.
- For processes of the type discussed here, the overestimation increases with the demand rate and the coarseness of the partitioning.
- Finer partitioning can mitigate the effects of nonstationarity on both the expected value and variance of the estimator.
- The bias function probably has little or no effect on system performance.

These findings, coupled with the inference from the Bitburg data discussed in Section 3 that there is some level of nonstationarity present in these demand processes, are consistent with:

- The association of larger values of observed VTMRs with high-demand items,
- The substantial differences between the distributions of VTMRs observed in the Bitburg data and those discussed by Crawford, and
- The wildness of the distribution of observed VTMRs around the power function used by AFMC to specify VTMRs to describe the probability distribution of the number of items of each type in re-supply.

One's intuition may be inappropriately shaped by the specific example of the ramp process discussed here. Such a high level of nonstationarity may seldom be seen in real-world parts demand processes; however, it clarifies the profound importance of explicit recognition of the roles of nonstationarity, data partitioning, and demand rates on the distributions of observed VTMRs. It seems fairly clear that, in general, we tend to overestimate the stochastic variability of these demand processes, especially those of high-demand items. In the discussion that follows, we point out the need to consider explicitly all the other sources of uncertainty that affect the demand process in our estimation of variance.

IMPROVED VARIANCE ESTIMATION

The variance of the demand process is one of the two most important elements of the forecasting problem. Our estimation of the variance may well have greater effects on the performance of the system than our estimation of the mean. Variance estimation tends to be a bit neglected relative to its importance, owing to the traditional logic of understanding the mean process before the variance of demands around the mean is addressed. The discussion of variability in Section 4 focuses attention on several separate components of the variance of interest in this discussion. The fundamental logical problem underlying variance estimation is to clarify our assumptions about nonstationarity and to explicate the difference between stochastic variability and forecast variance.

If we assume stationarity, then the current variance estimation procedure implicitly tries to estimate stochastic variability. It then uses this estimate as an estimate of forecast variance. If these demand processes really were stationary, then the current procedure would actually underestimate the forecast variance somewhat because the forecast variance should include both parameter variance and stochastic variability. The problem is, if there is nonstationarity present in these processes, then the current procedure overestimates stochastic variability because it fails to account for the nonstationarity. To make matters worse, the current procedure then uses this poor estimate of stochastic variability as an estimate of forecast variance. But with nonstationarity present, under an explicit assumption of nonstationarity, the forecast variance should include not only stochastic variability and parameter variance but should also account for the effects of the nonstationarity. Because it overestimates stochastic variability and fails to include other sources of variability, the current procedure enjoys the effects of errors in the opposite direction that don't necessarily cancel out; i.e., it isn't as bad an estimate of forecast variance as it might be. Interestingly, this is so quite by chance.

The appropriate variance to calculate is a function of the decision being made. Here we focus on the forecast variance associated with three-year procurement lead times. Decisions whose effects are felt over shorter planning horizons (e.g., annual repair planning) would be less vulnerable to forecast uncertainty owing to the reduced effects of nonstationarity across the shorter planning horizon. Once we have explicated precisely what forecast variance we are calculating, we can consider how to estimate the variance and how well it can be determined. Unfortunately, on the latter count we are in no better a posi-

tion with forecast variances than we were with process variances; these are highly variable processes and their variances are difficult to estimate.

We should also be clear about precisely what we mean by nonstationarity. In addition to nonstationarity in the mean, we could also have nonstationarity in the stochastic variability. As a matter of fact, there is every reason to believe that this is true. There is also every reason to believe that we will be unable to do a good job of estimating this nonstationarity. As we have shown in earlier sections, the variance of the VTMR estimator is large even for stationary processes. We have an even more difficult problem in the face of nonstationarity in the stochastic variability. The methods we propose here do not explicitly incorporate this nonstationarity.

As in any problem where we have inadequate information to estimate variances confidently, we must look for additional information to enhance the estimation process. The information might come from other parts or other relationships. The current system's method of estimating variances is to look at the relationship between means and VTMRs across many parts and then use the (presumably) better estimated means to specify the variances. There are other ways to improve our estimates by pooling information. Parts could be grouped together in various ways and their variances jointly estimated, as with "shrinkage" estimators, for example. More data could be obtained through more detailed data collection either by partitioning the data more finely in time or by using base-specific data. Finally, other, better estimated quantities could be used to develop more finely tuned estimators of variance along the lines of the present system's VTMR estimator.

We were unable to explore some of these methods because we did not have access to data partitioned more finely than by quarter, or to base-specific data. The effects of finer partitioning of the data are suggested by results presented above. We think this is a potentially fruitful area for improvement. Base-specific data would not only be useful for variance estimation but for stock level and asset allocation decisionmaking as well.

It is important to mention ideas we explored that did not pan out. One seductive possibility is to produce a model like the current system's nonlinear function for VTMR estimation that incorporates additional explanatory variables. In particular, there seems to be a logical disconnect in the current system. The current system would take two parts with radically different historical variability and predict the same variance if they had the same mean. An obvious fix for this dis-

connect is to develop a model for the variance that incorporates historical variability as well as historical means. The problem here is what to use as a measure of historical variability. We already know that the variance is poorly estimated by the VTMR estimator. Indeed, we can find a statistically significant relationship between historical variability and future variability even after controlling for the mean. But an estimator based on this relationship does not work well. The variability in the variance estimates is just too large.

With the lack of more finely partitioned data and our inability to incorporate part-specific variance estimates into the forecast variance, we are left with only one option. We must explore a mean-variance relationship such as the one used in the current system and see if we can improve it. There is reason to believe that we can improve on the current system's nonlinear function. Our hopes rest on the consequences of our explicit assumption of nonstationarity. If we consider nonstationarity, the current system is estimating the *wrong variance*. Hence if we apply the current system's general approach of using a nonlinear regression function to the three-year forecasts, we might improve the VTMR estimator.

To develop the improved variance-to-mean relationship we can take the three-year forecast errors and try to characterize their magnitude as a function of their forecast means. The method here is to make forecasts with whatever method one intends to use, calculate the three-year forecast errors, then fit a nonlinear function in a manner similar to that of the current system to forecast the variance (or VTMR).

Our best mean forecaster is the weighted regression forecaster for Group 1 and the current system's eight-quarter moving average for Groups 2 and 3. We applied these methods to our sample of 600 parts. We made forecasts for 10, 11, 12, and 13 quarters in the future and for each of those quarters computed the observed squared error of the forecast. This corresponds roughly to a three-year procurement lead time. We then performed a simple linear regression of the log of the squared forecast errors on the log of the forecast. This gives us a variance-to-mean relationship for our particular three-year forecasting method. This could be readily applied to other forecasts needed for planning and procurement. Indeed, this relationship should be reestimated each year. This is especially important in times of significant changes in funding levels, system structure, activity levels, or other factors that could affect future pipelines. It could also be tailored to item-specific procurement lead times. Note that the variance now includes some of the uncertainty stemming from nonstationarity.

The magnitude of this variance could well be affected by system restructuring.

Our regression yielded the relationship:

$$VTMR = 0.57 MEAN^{0.47},$$

constrained to be between 1.01 and 5.0.

Other explanatory variables might be useful in estimating VTMRs, for example, the QPA (as we observed in the Bitburg data); however, the QPA is specific to an item's application rather than to the item itself and is of limited utility at the aggregate, systemwide level where these estimates are being made.

The ultimate measure of the efficacy of improved demand and variance forecasting is the magnitude of its improvement in system performance. In the section that follows, we discuss our evaluations of these improved techniques.

7. EVALUATIONS OF THE IMPROVED DEMAND FORECASTING AND VARIANCE SPECIFICATION TECHNIQUES

As we pointed out above, 600 items were used to develop the improved demand forecasting and variance specification techniques, only 200 of which were high-demand (15 or more demands per quarter) items. We then tested these methods on all high-demand parts in the dataset, 4,215 items. The improvements in root mean squared error and mean absolute deviation achieved by using the improved techniques in contrast to the current system are reflected in Table 7.1. The improvement is impressive. (Also see Figures 5.3 and 5.4.)

Beyond evaluating the improved demand and variance forecasting techniques using ordinary statistical measures, it is important to understand how much they might improve the cost-effectiveness of the spares procurement mix over a realistically long planning horizon. We developed a system of software that replicates AFMC's spares requirements computation and central stock leveling system and evaluates the performance of the resulting spares stockage posture. We used this software system to evaluate the improved demand forecasting and variance specification techniques. The system is described in Figure 7.1 (except that the WRSK requirements portion of the software was not used in these evaluations). To evaluate these techniques, we used the March 1986 requirements database used by AFMC to compute spares and repair requirements and replicated the requirements computation, first with the current system's forecasting techniques and then with the improved techniques. We input the results to a replica of AFMC's central stock leveling system (D028), which allocates stock levels to bases and the depot, and added the war readiness spares the units are authorized, thereby estimating the

Table 7.1

**Percentage of Improvement in RMSE and
MAD of Improved Techniques over Current
System**

Measure	10-Quarter Horizon	13-Quarter Horizon
RMSE	48	38
MAD	51	45

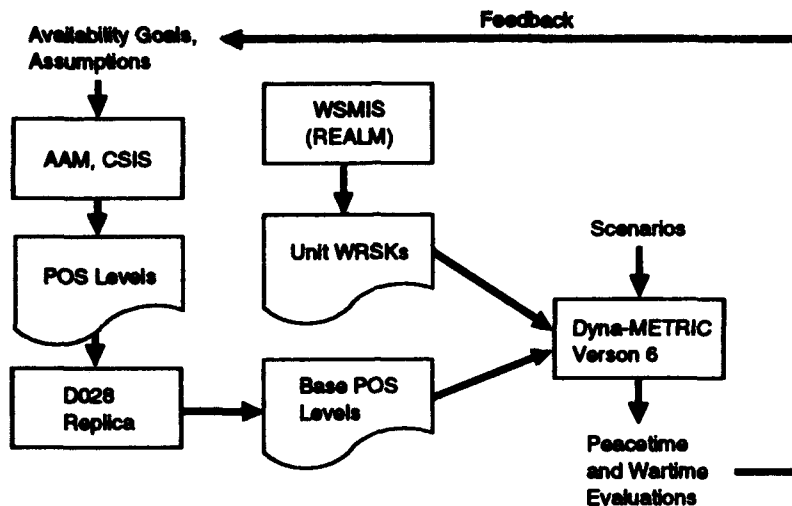


Figure 7.1—Schematic Representation of Evaluation System Software

anticipated stockage posture that would eventuate in the system roughly three years (actually 13 quarters) after the requirements computation. We then evaluated four stockage postures: (1) One anticipated to result from the current system using the same aircraft availability goals as used in the current system, (2) one anticipated to result from use of the improved techniques using the same specified goals, (3) one anticipated to result from the first approach coupled with roughly a \$240 million budget reduction, and (4) one anticipated to result from the improved techniques coupled with a budget roughly equal to that in case 3. In these evaluations, war readiness spares were not included. Using the case of the F-16 aircraft, we evaluated each of these stockage postures with an advanced capability assessment model, Dyna-METRIC Version 6, under several sets of assumptions that we will describe.

The evaluations of the four stockage postures were done with item characteristics drawn from the March 1990 database. The database contains eight quarters of past history of demands, NRTS actions, and so forth. Those data were used to evaluate the stockage postures anticipated to result from the two alternative forecasting methods in use with the March 1986 database. Thus the world eventuated differently than anticipated at the time the requirements computation was done, and the differences are explicitly accounted for in these

evaluations. The demand rates and variances actually observed from April 1988 through March 1990 were used for the evaluations except that variances were constrained to be less than or equal to 15. These data overlap the point in time, June 1989, that was an average lead time beyond the buy point for the FY87 budget. The results of the evaluations, shown in Tables 7.2 and 7.3, are disappointing in the sense that the effects of the improved techniques are largely masked by the much more dramatic effects of the management adaptations assumed to be in place in three of the scenarios.

In Tables 7.2 and 7.3, the case labeled "No cannibalization" represents the situation in which there is no consolidation of parts shortages (cannibalization) among aircraft. The second case, labeled "Full cannibalization" reflects the assumption in the evaluations of cannibalization of *all* parts shortages that increase aircraft availability. The third case adds another management adaptation, that of lateral supply. The fourth case adds a more responsive depot repair system, one that expedites the transportation, handling, and processing of components, reducing pipeline times from an average of 89 days to an average of 56 days for overseas bases and 48 days for CONUS bases.

Three points are worthy of note in Tables 7.2 and 7.3. The first is that the investment level and system performance that result from specifying the aircraft availability goals in the traditional way (Table 7.2) are superior with the improved methods because they deliver somewhat better performance with \$76 million less budgetary requirement. The second point is that for roughly equal budgets (Table 7.3), the improved techniques deliver generally better performance. The final point is that the performance with the improved methods and reduced budget is almost as good as that of the current system with an unreduced budget; however, the evaluations are somewhat difficult to interpret because all four of these cases are on such a flat part of the availability/cost curve; i.e., substantial budgetary changes produce relatively little effect on performance, suggesting that the stockage and budgetary requirements being computed by the current system are quite rich.

Given a different database (e.g., March 1987 or 1988), the outcome of this experiment might have been somewhat different, of course. What we have shown is that, *for any given budget level, the improved techniques are likely to deliver improved performance simply because they substantially reduce the expected forecasting error.* Forecasting error acts exactly like churn does in its effects on system performance. It is another realization of a future that we did not expect when we specified item parameters to the requirements computation.

Table 7.2
Cost and Performance with Traditional Availability Goals

Management Adaptations	Percentage of Aircraft Unavailable, Peacetime	
	Current System, \$3,709 Million	Improved Methods, \$3,633 Million
No cannibalization	74.9	71.7
Full cannibalization	33.0	32.3
Cannibalization, lateral supply	17.3	16.4
Cannibalization, quick, lateral supply	3.2	3.1

Table 7.3
Cost and Performance with Reduced Budgets

Management Adaptations	Percentage of Aircraft Unavailable, Peacetime	
	Current System, \$3,474 Million	Improved Methods, \$3,470 Million
No cannibalization	81.5	76.3
Full cannibalization	34.5	33.1
Cannibalization, lateral supply	19.0	17.2
Cannibalization, quick, lateral supply	3.5	3.6

In these evaluations, the improvements in variance specification and forecasting translate to improved system performance as well.

Note that we did not explicitly evaluate the improved variance specification technique independently. One would expect very little effect from moving to the use of the new variance specification simply because it is so much like the formula currently in use. The point should be made again, however, that what is important about the new technique is that it explicitly considers forecasting uncertainty and is sensitive to the length of the planning horizon. Thus its underlying logic is quite different from the logic of the current variance specification method. It is this difference that is important rather than the specific numerical values in the formulas.

It is also important to note that the budgetary values shown in these evaluations apply only to the first year of implementation. It is not clear what the savings in subsequent years would be relative to those in the first year.

8. CONCLUDING REMARKS

We conclude from this research that:

- A weighted regression forecasting technique would perform better than the eight-quarter moving average for high-demand items, that is, for items with 15 or more demands per quarter.
- A modification to the power function currently used by AFMC to specify the variance of the number of items in resupply would yield an improved mix of spares for a specified investment level.
- A NRTS forecaster that incorporates exponential weighting of past NRTS rates, coupled with the improved demand forecaster, would improve forecasts of NRTS actions that are important to depot repair planning.

These three improvements could yield substantial savings in spares costs for specified levels of system performance. In the spares requirements computation done with the March 1986 database, these changes, coupled with an investment level \$239 million less than the current system's, yielded roughly the same system performance.

We recommend implementation of the weighted regression forecaster with an α of 0.75 and a VTMR specification of $0.57 \text{ MEAN}^{0.47}$, constrained to be in the range of 1.01 to 5.0. The 0.75 weighting factor is also recommended for the improved NRTS rate forecaster.

In the D035C data system, supply transactions are reported daily; therefore, there may be some way to move to the use of a finer partitioning of demand data than is used in the current system. This would make the VTMR estimator considerably less vulnerable to nonstationarity. For some purposes, quarterly historical demand data may be sufficient, but for purposes of variance estimation, more finely partitioned data would clearly be more helpful.

In retrospect, we view the recommendations above as only one important outcome of this research. The explicit recognition and treatment of nonstationarity in these parts demand processes are, perhaps, more important ideas in the longer-term scheme of things. In future research in this area, this consideration is especially important. Any future improvements in spares demand modeling and forecasting will probably depend on these ideas.

This research leaves unresolved one of the most important problems complicating spares and repair requirements estimation: the estimation of wartime demand rates. We do not intend to imply by this that we think we have resolved all other problems, but we do believe that the results reported here represent a significant improvement to the approach to demand forecasting taken by the current system.

We recommend the following topics for future research:

- Extensions of the analysis of base-level demand processes to additional bases and weapon systems,
- Exploration of the use of empirical Bayes estimators in the AFMC Central Leveling System (D028), and
- Examination of alternative methods for modeling the demands for low-demand items.

As pointed out in Section 3, our observations about base-level demand processes were based on the Bitburg data. The analyses of these data need to be extended to other bases and weapon systems. The motivation for these extensions is to ensure that we are modeling base-level demand processes sensibly in decision support systems used centrally.

It has long been hypothesized by Crawford and others, including the current authors, that neither base-specific nor worldwide demand rates are the best estimators of base-specific future demands. Preliminary evaluations of empirical Bayes estimators that adjust base-specific demand rates by worldwide means tend to reinforce this hypothesis; however, the sample size of these evaluations was too small to be conclusive, although they served to strengthen our intuition that the hypothesis is a sensible one. Unfortunately, the ability to evaluate this proposed approach would depend on a data-collection effort that could be quite extensive and time consuming. The problem is that data describing base-specific demands over time are needed to support the evaluations. This requirement could conceivably be supported by NRTS actions reported in the Maintenance Data Collection System, although base supply transaction data would clearly be better for the purpose. We have been led to believe that such data could be collected through the D024 system, but the effort would require at least four years of transaction data from several bases. Although such a data-collection effort seems quite ambitious, the data could serve several interests, including central stock leveling techniques.

One of the problems with the approach to the weighted regression technique discussed here is that it assumes normal distributions of

demands. While this may be a reasonable approximation for high-demand items, it is flawed as a model for lower-demand items, especially for items with very low demands. We regret not being able to suggest additional avenues of exploration that hold promise of success with low-demand items, but hypotheses should be encouraged and evaluated even though low-demand items typically have less impact on system performance.

Another important problem involved with modeling demands for low-demand items is that past demand data provide too little information about the demand processes of individual items. Thus a second idea emerges: to borrow strength from the demand history of many low-demand items to enhance our ability to model demands for individual items. This idea seems as reasonable now as when it first emerged from earlier RAND research. Unfortunately, it was never pursued with sufficient thoroughness to lead to implementation. The idea of pooling data across items has Bayesian roots, of course, as does the Kalman filter. They are compatible in philosophy and could be combined with empirical Bayes approaches to central leveling. This would constitute a unified Bayesian approach to demand modeling to support the central allocation of stock levels as well as spares and repair requirements estimation.

Appendix A

RELATIONSHIP OF THE KALMAN FILTER TO OTHER FORECASTING METHODS

One useful characteristic of Kalman filter models is the way they generalize other well known time series and forecasting methods. Unfortunately, this has caused some confusion in the literature. It is not correct to claim that Kalman filtering is "just exponential smoothing," as is sometimes heard. Exponential smoothing is a special case of Kalman filtering but Kalman filtering incorporates a much richer collection of models. In the discussion that follows, we demonstrate how some well-known statistical models can be written as Kalman filter models. Refer to Ref. 63 for a more complete exposition.

I. A first order autoregressive process in the state equation.

The state equation is

$$\beta_t - \beta = \phi(\beta_{t-1} - \beta) + w_t, \quad 0 \leq \phi \leq 1.$$

The measurement equation is

$$Y_t = X\beta_t + v_t.$$

Some useful facts:

1. $\text{corr}(\beta_t, \beta_{t-j}) = \phi^j$,
2. If $\phi = 0$ this is a random effects model,
3. If $\phi = 1$ this is the "simple" Kalman filter model.

II. A first order vector autoregressive process in the state equation.

Replace β and β_t with vectors and ϕ with a square matrix in example I.

III. A vector autoregressive moving average process in the state equation. β and w are vectors and ϕ and θ are square matrices.

The state equation is

$$\begin{aligned}\beta_t - \beta &= \phi_1(\beta_{t-1} - \beta) - \dots - \phi_p(\beta_{t-p} - \beta) \\ &= w_t - \theta_1 w_{t-1} - \dots - \theta_q w_{t-q},\end{aligned}$$

where p is the number of observations underlying the moving average and q is the number of steps in the autoregression.

IV. Example I is equivalent to certain linear models.

The state equation is

$$\beta_t - \beta = \phi(\beta_{t-1} - \beta) + w_t.$$

The measurement equation is

$$Y_t = X\beta_t + v_t.$$

Now manipulate the measurement equation

$$Y_t = (X\beta + (X\beta_t - X\beta)) + v_t.$$

Let $(X\beta_t - X\beta) + v_t = a_t$. So $Y_t = X\beta + a_t$, where a_t is heteroskedastic and autocorrelated. If the structure and values of the elements of a_t 's variance-covariance matrix are known, the parameters can be found using generalized least squares.

V. The simple Kalman filter model is similar to an exponential smoother.

The state equation is

$$\theta_t = \theta_{t-1} + w_t.$$

The measurement equation is

$$Y_t = \theta_t + v_t.$$

This model corresponds to a class of ARIMA(0,1,1) models from Box and Jenkins [64].

If we set $V_t = 2$ and $W_t = 1$ we can use the update equation

$$\beta_t = \beta_{t-1} + \frac{1}{2}(Y_t - \beta_{t-1})$$

to solve for θ_t as a function of the observations and initial conditions

$$\beta_t = \sum_{j=0}^{t-1} \left(\frac{1}{2}\right)^{j+1} Y_{t-j} + \left(\frac{1}{2}\right)^t \beta_0$$

This reveals the filter to be a simple exponential smoother with a starting value incorporated. Of course, the Kalman filter provides standard errors, an intuitive reason for the smoothing parameter, and obvious ways to extend the model if the fit is inadequate.

Appendix B

THE FUNCTIONAL FORM OF THE MEAN DEMAND PROCESS

Although the simple multiplicative function for the mean demand rate that the current system uses is incorrect, finding the best alternative is a difficult problem. The process of selecting a model based on the data is itself not a well understood problem even for stationary, normal models. The approach we take here is to conduct an automated model search on the sample of parts used in Section 4.

A DATA-ANALYTIC APPROACH TO DETERMINING THE FUNCTIONAL FORM

We first examine a linear regression model for the mean demand as a function of flying hours, flying hours in the previous period, and demands in the previous period. Intuitively these quantities reflect three sources of demand. Flying hours reflect current stress on the aircraft. Flying hours from the previous period reflect the delayed effect of earlier stresses. Including demands from the previous period is an effort to capture stresses unrelated to flying hours with a duration longer than one quarter. The idea of using lagged demand is that there may be real, unmodeled stresses on the system reflected by previous demand. A fourth source of demand, a background level of demand unrelated to use, is reflected by the inclusion of a constant in the model. In this section we refer to these explanatory variables as X , to have a generic label.

We consider three possible ways to modify this basic model: (1) selecting a subset of these explanatory variables, (2) transforming the explanatory variables, and (3) transforming the demands. Unfortunately, the order in which we consider these possible modifications may have an effect on our conclusions. For example, if we decide to transform the demands by taking the logarithm, we may select different explanatory variables than if we had not transformed. We will use a model search strategy that should yield a model with good predictive quality [65]. That strategy has four steps: (1) select the explanatory variables, (2) examine the data for potential outliers, (3) transform the demands if needed, and (4) transform the explanatory variables if needed.

In the discussion that follows, we explain the automated procedures used for each of these stages, present the results of this strategy for model selection, and consider the plausibility of these results and other, related models.

DESCRIPTIONS OF THE COMPONENT PROCEDURES

Mallow's C_p

Mallow's C_p is a criterion-based variable selection method [66]. The idea is to minimize the mean squared error of prediction, J_p , where J_p is given by

$$J_p = \frac{1}{\sigma^2} \sum_{i=1}^n \text{mse}(\hat{Y}_i) ,$$

where n is the number of observations.

Of course, there are some unknown quantities here. As an estimate of J_p , Mallows [66] proposed

$$C_p = \frac{\text{RSS}_p}{\hat{\sigma}^2} + 2p - n .$$

RSS_p is the residual sum of squares from a model with p explanatory variables, and $\hat{\sigma}^2$ is the error estimate from the full model. The "best" model is the one with the smallest C_p . This method of variable selection is intended to produce models that predict well. The details of the computation can be found in Ref. 67, pp. 215-217.

Outlier Rejection Using the Bonferroni Inequality

This outlier test is quite intuitive. The regression is refit with the questionable point omitted. Under the usual regression assumptions the difference between the omitted point and the refitted line can be compared using the standard error for prediction based on the refitted equation. The actual calculation is as follows:

First, the residuals are calculated for each case by fitting the model with the case deleted, and the t ratio is formed to test the hypothesis that the residual is zero. This calculation can be done by using the relationship

$$t_i = \frac{e_i}{\sigma_{(i)} \sqrt{1 - h_i}} .$$

In this expression, e_i represents the raw residual for the i th case and h_i is the diagonal of the "hat" matrix, $X(X'X)^{-1}X'$. The t_i are the externally studentized residuals, although this terminology is not universal. $\sigma_{(i)}$ is estimated from the regression with case i omitted. All of these t_i are t -distributed with $n - p - 1$ degrees of freedom. The t_i with the largest absolute value is then tested to see if it is too large. Since we are picking the largest of n residuals, we need to get a critical value for the largest absolute value of n draws from the student's t distribution. Snedecor and Cochran [68] suggest testing the largest residual at the level α/n against the t distribution with $n - p - 1$ degrees of freedom. This Bonferroni procedure guarantees a test of no more than size α .

It is not actually necessary to recompute the regression with each observation omitted. There are more convenient formulas. All of this material is covered in Ref. 67, pp. 114-117.

Box-Cox Procedure

The Box-Cox procedure [69] selects a transformation of D automatically from a collection of power transformations. The Box-Cox model assumes there is a power transformation, indexed by the power λ , such that the usual linear model assumptions are satisfied, i.e.,

$$D^\lambda = X\beta + \epsilon, \text{ where } \epsilon \text{ is } N(0, \sigma^2) .$$

In this implementation, the likelihood function is calculated for λ in the set $\{-2, -1, -0.5, 0, 0.5, 1, 2\}$, where the power zero transformation corresponds to taking the logarithm. The value that maximizes the likelihood on this set is the selected transformation. Note that no test is done to see if a transformation is required.

This procedure requires D to be positive. If any value is zero, all D values are shifted so that the smallest value is 1.0.

Box-Tidwell Procedure

The Box-Tidwell procedure [62] provides an automated method of transforming the X s. Consider the model for the transformation of X_1 :

$$Y = \beta_0 + \beta_1 x_1^\alpha + \sum_{j=2}^p \beta_j x_j + \varepsilon .$$

Expanding x_1^α in a Taylor's series around 1.0 and deleting higher order terms yields

$$x_1^\alpha = x_1 + (\alpha - 1)x_1 \ln(x_1) .$$

This is substituted into the equation

$$Y = \beta_0 + \sum_{j=1}^p \beta_j x_j + \eta x_1 \ln(x_1) + \varepsilon ,$$

where a test of the hypothesis $\eta = \beta_1(\alpha_1 - 1) = 0$ tests the need to transform. This test is performed to determine whether a transformation is required. If η is significant, a transformation is performed. The indicated transformation is $\eta/\beta_1 + 1$. Since this might not be a very tidy value, the closest value in the set $\{-2, -1, -0.5, 0, 0.5, 1, 2\}$ was used. Each X variable is considered for transformation with all of the other X variables in their original scale. This ensures that the transformations for the Xs are not affected by the order in which they are considered.

THE RESULTS OF THE MODEL SEARCH

In this discussion we describe these four model selection steps and report on the results of applying them to the 600-part sample described in Section 4.

The first procedure applied was Mallows's C_p variable selection method. The various subsets and their prevalence in the three groups are shown in Table B.1. In the table, a zero indicates the variable is not in the model, and a one indicates that it is.

Overall, the most frequently selected model is one with only the previous period's demands included, although the flying hour and previous demand model is the most frequently selected for Group 1, the highest demand group. The flying-hour-only model ties for most frequent in group 2. It is always unsatisfying to use a model that has lagged demands as an explanatory variable. This is tantamount to an

Table B.1
Results of the Variable Selection Search

Explanatory Variables in the Model			Group 1	Group 2	Group 3
fh	fh ₋₁	D ₋₁			
0	0	0	14	39	48
0	0	1	42	47	50
0	1	0	13	21	29
1	0	0	29	47	34
1	0	1	49	14	11
1	1	0	7	7	11
1	1	1	16	4	5

admission that we don't really understand the process. It is also difficult to use this type of model to estimate counter-factuals. Questions of the form, "What if we fly more?" often need to be addressed. Despite these caveats, the results in the Table B.2 suggest a model with both flying hours and lagged demands.

It is worth noting the frequency with which outliers were rejected by this automated process. Table B.2 gives the counts.

These rejections could be the result of any number of problems with the data or the modeling. It is useful to set observations aside in this manner in an effort to capture the underlying structure of the data. We do not advocate discarding outliers during our predictive evaluation of methods. It is of little solace to know that the 30 important parts you needed but did not have were merely "statistical outliers."

After variable selection we used the Box-Cox procedure to explore the need for a transformation of the demands. The results of this examination are summarized in Table B.3.

Table B.2
Fraction of Outlier Rejections

Group 1	Group 2	Group 3
37/200	60/200	57/200

Table B.3
Indicated Power Transformation

Group	-2	-1	-0.5	0	0.5	1	2
1	0	2	10	26	88	66	8
2	13	10	18	68	80	11	0
3	13	14	22	58	72	21	0

The square root transformation is the most commonly indicated transformation in each of the three groups. This unanimity of results is encouraging.

The final model improvement that we will consider is transformation of the explanatory variables. Table B.4 summarizes these results.

The "Omit" category indicates that the model selection did not include that variable in the final model. These results indicate no need to transform the explanatory variables.

CRITIQUE AND SUMMARY

The variable selection portion of this model selection exercise occurred before the transformation of the demands was explored. Since the variable selection was performed on the untransformed values, there is some uncertainty as to whether the results would hold up if the variable selection had been done with the square root of demands.

Table B.4
Transformations of the Explanatory Variables

Group	Variable	Omit	Indicated Transformation						
			-2	-1	-0.5	0	0.5	1	2
1	fh	99	5	1	2	2	4	79	8
1	fh ₋₁	134	8	0	0	3	1	52	2
1	D ₋₁	63	5	5	8	8	1	108	2
2	fh	128	2	1	0	2	1	62	4
2	fh ₋₁	147	0	2	2	2	1	39	7
2	D ₋₁	114	6	5	8	0	0	65	2
3	fh	139	2	0	3	5	4	45	2
3	fh ₋₁	143	0	1	4	1	1	41	9
3	D ₋₁	122	5	3	0	8	0	61	1

Rerunning this analysis with the demands transformed before beginning leaves the variable selection and explanatory variable transformation results virtually unchanged.

The strong indication that the proper transformation of demands is the square root is reassuring. This is a standard transformation for discrete data of this type. The inclusion of demands from previous periods in the model is discouraging and may be an indication that the model omits important explanatory variables. However, these are the explanatory variables that are available to us and we must live with them. Lagged demands are unsatisfying predictors but may be useful predictors.

The deletion of lagged flying hours from the models should be taken with a grain of salt as well. Typically, flying hours do not change much from quarter to quarter. This strong correlation between adjacent quarters could easily be responsible for the deletion of lagged flying hours from the model. If scenarios with erratic flying hour programs are considered, more examination of the predictive usefulness of this variable would be in order.

REFERENCES

1. Lippiatt, Thomas F., *Variability in the Budget Forecasts for Depot-Level Component Repair*, RAND, N-2930-P&L, January 1991.
2. Crawford, G. B., *Variability in the Demands for Aircraft Spare Parts: Its Magnitude and Implications*, RAND, R-3318-AF, January 1988.
3. Geisler, M. A., B. B. Brown, and O. M. Hixon, *Analysis of B-47 Consumption Data and Activity*, RAND, RM-1288, July 1954.
4. Brown, B. B., and M. A. Geisler, *Analysis of the Demand Patterns for B-47 Airframe Parts at Air Base Level*, RAND, RM-1297, July 1954.
5. Geisler, M. A., and J. W. T. Youngs, *Predictability of Demand for B-47 Airframe Spare Items*, RAND, RM-1300, August 1954.
6. Brown, B. B., and M. A. Geisler, *Comparison of Demand for Spare Aircraft Items by B-50D Aircraft (Project Red Head) and B-47 Aircraft Deployed to English Bases*, RAND, RM-1333, September 1954.
7. Karr, H. W., *Analysis of B-47 AOCP Experience*, RAND, RM-1340, September 1954.
8. Youngs, J. W. T., M. A. Geisler, and A. R. Mirkovich, *Confidence Intervals for Poisson Parameters in Logistics Research*, RAND, RM-1357, September 1954.
9. Youngs, J. W. T., M. A. Geisler, and B. B. Brown, *The Prediction of Demand for Aircraft Spare Parts Using the Method of Conditional Probabilities*, RAND, RM-1413, January 1955.
10. Greenwood, M., and G. U. Yule, "An Inquiry into the Nature of Frequency-Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents," *Journal of the Royal Statistical Society*, Vol. 83 (1920), p. 255.
11. Geisler, M. A., *Analysis of Base Stockage Policies*, RAND, RM-1431, February 1955.
12. Geisler, M. A., and A. R. Mirkovich, *Analysis of the Flying Activity and Spare Parts Demand of F-86D Aircraft at Perrin Air*

Force Base, September 1, 1953–February 28, 1954, RAND, RM-1456, April 1955.

13. Geisler, M. A., and A. R. Mirkovich, *Analysis of Worldwide Data on Aircraft Spare Parts as to Unit Cost, Quantity and Value Issued, and Inventory Value*, RAND, RM-1481, May 1955.
14. Hamburger, William, *Simple Distribution Functions for Inventory Control*, RAND, RM-1539, August 1955.
15. Karr, H. W., M. A. Geisler, and B. B. Brown, *A Preferred Method for Designing a Flyaway Kit*, RAND, RM-1490, May 1955.
16. Hamburger, William, *The Effectiveness of Alternative Flyaway Kits*, RAND, RM-1541, August 1955.
17. Clark, A. J., *A Technique for Optimal Distribution of Available Stocks to Bases*, RAND, RM-1621, January 1956.
18. Berman, E. B., *A Model of the Procurement-Repair Decision for a Spare Item*, RAND, RM-1519, July 1955.
19. Brown, B. B., *Characteristics of Demand for Aircraft Spare Parts*, RAND, R-292, July 1956.
20. Cohen, I. K., John B. Abell, and Thomas F. Lippiatt, *Coupling Logistics to Operations to Meet Uncertainty and the Threat (CLOUT): An Overview*, RAND, R-3979-AF, 1991.
21. Goldman, T. A., *Relationships between Program Elements and System Demand for Airframe Spare Parts*, RAND, RM-1858, January 1957.
22. Goldman, T. A., *A Priori Demand Prediction—A Case Study of B-52 Airframe Parts*, RAND, RM-2088, January 1958.
23. Goldman, T. A., *System Demand for Airframe Parts During Weapon-System Phase-In*, RAND, RM-2381, May 1959 (FOUO).
24. McGlothlin, W. H., et al., *The Simulated Aircraft and Its Failure Model in LP-I*, RAND, RM-2177, May 1958.
25. McGlothlin, W. H., and R. Radner, *The Use of Bayesian Techniques for Predicting Spare-Parts Demand*, RAND, RM-2536, March 1960.
26. McGlothlin, W. H., and E. E. Bean, *Application of the Bayes Technique to Spare-Parts Demand Prediction*, RAND, RM-2701, January 1961.

27. Astrachan, M., B. B. Brown, and J. W. Houghten, *A Comparative Study of Prediction Techniques*, RAND, RM-2811-PR, December 1961 (FOUO).
28. Astrachan, M., and A. S. Cahn (eds.), *Proceedings of RAND's Demand Prediction Conference, January 25-26, 1962*, RAND, RM-3358-PR, January 1963.
29. Campbell, H. S., *The Relationship of Resource Demands to Air-base Operations*, RAND, RM-3428-PR, January 1963 (FOUO).
30. Feeney, G. J., J. W. Petersen, and C. C. Sherbrooke, *An Aggregate Base Stockage Policy for Recoverable Spare Parts*, RAND, RM-3644-PR, June 1963.
31. McGlothlin, W. H., *Development of Bayesian Parameters for Spare-Parts Demand Prediction*, RAND, RM-3699-PR, July 1963.
32. Fishman, G. S., *A Statistical Description of Demand Patterns Evolving through Time*, RAND, RM-3704-PR, August 1963.
33. Astrachan, M., and C. C. Sherbrooke, *An Empirical Test of Exponential Smoothing*, RAND, RM-3938-PR, March 1964.
34. Feeney, G. J., and C. C. Sherbrooke, *Systems Analysis and Supply Management*, RAND, RM-4054-PR, April 1964.
35. Feeney, G. J., and C. C. Sherbrooke, *The (s-1,s) Inventory Policy under Compound Poisson Demand: A Theory of Recoverable Item Stockage*, RAND, RM-4176-1-PR, March 1966, originally published with the same title as RM-4176-PR, September 1964.
36. Palm, C., "Analysis of the Erlang Traffic Formula for Busy-Signal Arrangements," *Ericsson Technics*, No. 5, 1938, pp. 39-58.
37. Feeney, G. J., and C. C. Sherbrooke, *An Objective Bayes Approach for Inventory Decisions*, RAND, RM-4362-PR, March 1965.
38. Feeney, G. J., and C. C. Sherbrooke, *A System Approach to Base Stockage of Recoverable Items*, RAND, RM-4720-PR, December 1965.
39. Press, S. J., *Statistical Estimation by the Empirical Bayes Method: Some Extensions and Logistical Applications*, RAND, RM-4442-PR, June 1965.
40. Robbins, H., "An Empirical Bayes Approach to Statistics," *Proc. of the Third Berkeley Symposium on Math. Stat. and Prob.*,

J. Neyman (ed.), Vol. 1, University of California Press, Berkeley, 1956, pp. 157-163.

41. Campbell, H. S., J. Y. Lu, and G. A. Michels, *War Readiness Materiel (WRM) Program for Dispersal Sites: A Systems Approach to Recoverable Spares Stockage (U)*, RAND, RM-4765-PR, January 1966, SECRET.
42. Houghten, J. W., *Matching the AFLC Requirements Computation Policy to the Official Base Stockage Policy*, RAND, RM-4301-PR, July 1966.
43. Sherbrooke, C. C., *Discrete Compound Poisson Processes and Tables of the Geometric Poisson Distribution*, RAND, RM-4831-PR, July 1966.
44. Sherbrooke, C. C., *METRIC: A Multi-Echelon Technique for Recoverable Item Control*, RAND, RM-5078-PR, November 1966.
45. Little, J. D. C., "A Proof for the Queuing Formula $L = \lambda W$," *Operations Research* 9, No. 3, May 1961.
46. Sherbrooke, C. C., *A Management Perspective on METRIC—Multi-Echelon Technique for Recoverable Item Control*, RAND, RM-5078/1-PR, January 1968.
47. Lu, J. Y., and R. B. S. Brooks, *An Aggregate Stockage Policy for EOQ Items at Base Level*, RAND, RM-5678-PR, June 1968.
48. Fishman, G. S., *Improved Forecasting Methods for Item Management and Maintenance Manpower Requirements*, RAND, RM-6010-PR, June 1969.
49. Muckstadt, J. A., *Comparative Adequacy of Steady-State Versus Dynamic Models for Calculating Stockage Requirements*, RAND, R-2636-AF, November 1980.
50. Hillestad, R. J., *Dyna-METRIC: Dynamic Multi-Echelon Technique for Recoverable Item Control*, RAND, R-2785-AF, March 1982.
51. Isaacson, Karen E., et al., *Dyna-METRIC Version 4: Modeling Worldwide Logistics Support of Aircraft Components*, RAND, R-3389-AF, May 1988.
52. Pyles, R. A., *The Dyna-METRIC Readiness Assessment Model: Motivation, Capabilities, and Use*, RAND, R-2886-AF, June 1984.

53. Isaacson, Karen E., and Patricia M. Boren, *Dyna-METRIC Version 5: A Capability Assessment Model Including Constrained Repair and Management Adaptations*, RAND, R-3612-AF, August 1988.
54. Crawford, G. B., *Palm's Theorem for Nonstationary Processes*, RAND, R-2750-RC, October 1981.
55. Carrillo, Manuel J., *Generalizations of Palm's Theorem and Dyna-METRIC's Demand and Pipeline Variability*, RAND, R-3698-AF, January 1989.
56. Hodges, James S., with Raymond A. Pyles, *Onward Through the Fog: Uncertainty and Management Adaptation in Systems Analysis*, RAND, R-3760-AF/A/OSD, July 1990.
57. Abell, John B., et al., *DRIVE (Distribution and Repair in Variable Environments): Enhancing the Responsiveness of Depot Repair*, RAND, R-3888-AF, 1992.
58. Abell, John B., et al., *Estimating Requirements for Aircraft Recoverable Spares and Depot Repair*, RAND, R-4210-AF, 1993.
59. Stevens, R. J., and J. M. Hill, *Estimating the Variance-to-Mean Ratio for Recoverable Items in the ALS Marginal Analysis Algorithms*, Working Paper Number 49, Headquarters, AFLC, Wright-Patterson AFB, Ohio, March 1973.
60. Hodges, James S., *Modeling the Demand for Spare Parts: Estimating the Variance-to-Mean Ratio and Other Issues*, RAND, N-2086-AF, May 1985.
61. Kalman, R. E., and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," *Journal of Basic Engineering*, Vol. 83, pp. 95-108, 1961.
62. Box, G. E. P. and Tidwell, P. W., "Transformations of the Independent Variables," *Technometrics*, 4, 531-550, 1962.
63. West, Mike, and Jeff Harrison, *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, New York, 1989.
64. Box, G. E. P., and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, 1970.
65. Adams, John L., *Evaluating Regression Strategies*, Ph.D. Thesis, University of Minnesota, 1990.
66. Mallows, C. L., "Some Comments on C_p ," *Technometrics*, Vol. 15, pp. 661-675, 1973.

67. Weisberg, S., *Applied Linear Regression*, Second Edition, John Wiley & Sons, New York, 1985.
68. Snedecor, G. W., and W. G. Cochran, *Statistical Methods*, Sixth Edition, The Iowa State University Press, 1978.
69. Box, G. E. P., and D. R. Cox, "An Analysis of Transformations," *Journal of the Royal Statistical Society, Ser. B.* 26, 211-252, 1964.
70. Isaacson, Karen E., and Patricia M. Boren, *Dyna-METRIC Version 6: An Advanced Capability Assessment Model*, RAND, R-4214-AF, 1993.